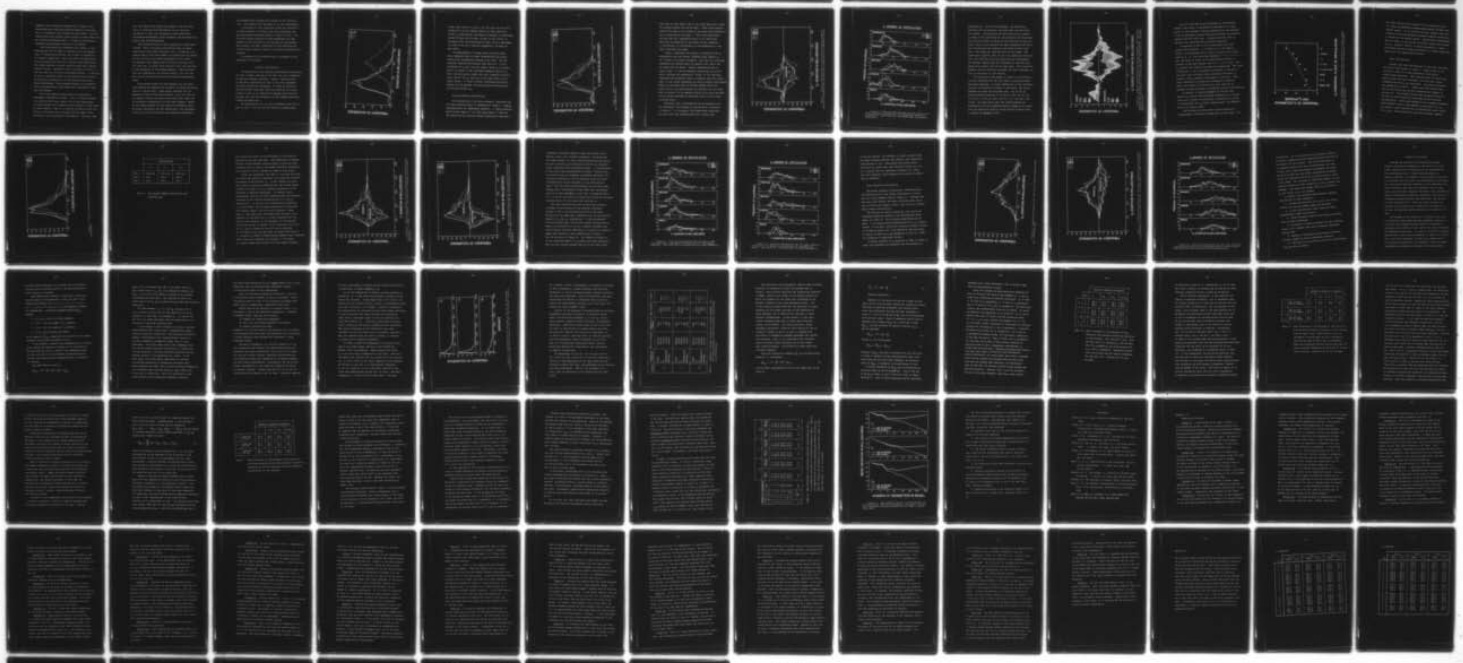
AD-A037 256 SPEECH COMMUNICATIONS RESEARCH LAB INC SANTA BARBARA CALIF F/G 5/7
TEMPORAL INTERRELATIONS IN SELECTED ENGLISH CVC UTTERANCES.(U)
MAY 76 R H FERTIG F44620-69-C-0078
UNCLASSIFIED SCRL-MONOGRAPH-12 NL

1 OF 1
AD A037256



END

DATE FILMED
4-77

SCRL MONOGRAPH NO. 12

12

MAY 1976

TEMPORAL INTERRELATIONS IN SELECTED ENGLISH CVC UTTERANCES

RALPH H. FERTIG

SPEECH COMMUNICATIONS RESEARCH LABORATORY, INC.

SANTA BARBARA, CALIFORNIA

(14) SCRL-Monograph Number-12

SPEECH COMMUNICATIONS RESEARCH LABORATORY, INC.

SANTA BARBARA, CALIFORNIA

(6) Temporal Interrelations in Selected English CVC Utterances,

(9) Interim rept.,

(11) May 1976

(12) 88 p.

(10) Ralph H. Fertig

(15) F44620-69-C-0078,
F44620-74-C-0034

Acknowledgement

This monograph is the result of a long research project which commenced under the auspices of Dr. Gordon E. Peterson, founder and first director of the Speech Communications Research Laboratory. His steadfast dedication toward the field of speech research in general combined with a specific concern toward this project has provided the basic impetus from which the monograph has grown.

Dr. Peterson's successor, Dr. June Shoup Hummel, has continued to provide support and assistance whenever needed. Grateful appreciation is further acknowledged to Dr. David J. Broad for his contribution to the work; his cogent observations and suggestions have been freely and generously given.

This research and publication has been supported by the Directorate of Mathematical and Information Sciences of the United States Air Force Office of Scientific Research under contracts F44620-69-C-0078 and F44620-74-C-0034. They have also been supported by the Office of Naval Research under contracts N00014-67-C-0118 and N00014-75-C-0483.

Temporal Interrelations in Selected
English CVC Utterances

by

Ralph H. Fertig

Abstract

The segment duration in a corpus of 1728 CVC utterances containing the vowel /ɪ/ by a single speaker were measured and analyzed. Measurement of sound spectrograms was facilitated by a computer-assisted television display. Various distributions of the durations under different conditions reveal a number of effects on segment durations attributable to consonant identity, consonant voicing characteristics and manners of articulation. Analyses of variance show that an interaction effect between the two consonants is statistically not significant for the durations of either consonant or of the vowel. Various models for describing and predicting the durations are given together with a discussion of their associated errors. The segmentation criteria are given in some detail as an appendix. A second appendix contains basic means and standard deviations derived from the data.

Table of Contents

Temporal Interrelations in Selected English CVC Utterances

Introduction

The purpose of this study is to investigate some of the interrelationships among segment durations in a limited body of CVC utterances. The speech data consist of CVC utterances in which the vowel is the American English /ɪ/ phoneme and the preceding and following consonants, designated C_1 and C_2, respectively, range independently over 23 English phonemes /w,l,r,j,ʍ, f,v,θ,ð,h,s,z,ʃ,ʒ,p,b,t,d,k,g,m,n,ŋ/, and the condition of silence /#/. This set of 576 CVC monosyllables was recorded by a phonetically-trained native American male in three different orders on three separate days, yielding a body of 1728 utterances.

The vowel /ɪ/ was specifically selected for two reasons. First, it was chosen because of its variable second and third formants, deemed to be important for a related study on formant variation under consonantal influence (BROAD AND FERTIG, 1970). Second, /ɪ/ occurs in a larger proportion of possible English CVC triples

than any other English vowel. Out of the 576 possible CɪC triples, 90 can occur only internally within individual words, 56 can occur only across word boundaries, 286 can occur either way, and the remaining 144 do not occur at all (SHOUP, 1964). This means that 75% of the possible triples can occur in spoken English. An English vowel other than /ɪ/ can occur in more than a third of the remaining 25% of the possible contexts. The intentional effect, therefore, was to maximize the naturalness of the spoken triples. The speaker, indeed, felt that there was no difficulty in producing the 84% of the CVC triples which occur for /ɪ/ or another vowel, and that, with a little forethought, relative naturalness could be achieved on the remaining utterances as well.

There are two reasons for utilizing CVC syllables spoken in isolation rather than data collected from continuous speech. First, a corpus including all possible combinations permits the application of more powerful statistical tools in analysis of the data. Second, the use of the silence control on the outside of the triples eliminates effects from adjacent sounds.

Collecting such a large mass of data on only one vowel rather than spreading the research over several vowels has had as a goal a truer understanding of the statistical variability of the durations. It has resulted

in a view of many subtle relationships which require the thorough corroboration provided. The application of powerful statistical methods to large quantities of data such as these has not, to the knowledge of the author, been previously accomplished. It is hoped that in the future this study will be expanded to other vowels, additional speakers, and further phoneme combinations.

Technique

Acquisition of quantitative data from the utterances involved tape recording the sounds, making sound spectrograms, and measuring these spectrograms with a computer-controlled television reading system. A detailed description of the procedure follows.

The three sets of 576 CVC triples were recorded in a sound-absorbent recording room through an Altec 633-A microphone onto one of the two channels of an Ampex 350-2 tape recorder. Simultaneously, a 200 Hz calibration pulse train was recorded on the second channel. After completing each of the utterance lists, the speaker listened to the set and then re-recorded those triples which he felt did not represent his normal relaxed speech. This was repeated until the entire set was satisfactory to him.

Broad-band sound spectrograms were made of all 1728 utterances, with narrow-band calibration spectrograms of the 200 Hz tone being made once at the beginning and then just after each ninth utterance. A Bell Telephone Laboratories' sound spectrograph (KOENIG, DUNN, AND LACY, 1946) was employed. The calibration spectrograms were then measured with a standard template and dividers at the fifteenth harmonic (3000 Hz). The calibration measurements show two general tendencies: 1) a slight but evident drift over an entire recorded set, construed as a variation in tape speed during recording; and 2) a daily rise of calibration values, attributed to the rise in ambient temperature during spectrograph operation on a given day. The group of calibration measurements was next used to determine adjustment factors for each member of each set of nine utterance spectrograms through linear interpolation between the two closest calibration values. In this manner, an approximate representation of the frequency (and, thereby, duration) shift could be identified with each individual utterance.

Prior to spectrogram measurement, decisions were required on what to measure. In addition to the necessary three segment durations from each triple, information on the vowel formants was desired. The complex patterns of the formants, however, virtually preclude simple measurements. The following observations were made from vowel

formant movements in a pilot study: 1) when both the second and third formants rise to maxima (or fall to minima), they frequently do so at considerably different locations in time; 2) the formants often assume S-shaped curves; 3) formant maxima and minima tend to be located substantially prior to the vowel midpoint; and 4) in many instances, the formants rise or fall monotonically, with or without significant slope changes. Due to the appearance of these and other complex formant phenomena, it was decided that formant measurements had to be taken at multiple points in time. Accordingly, each vowel segment was divided into ten sections of equal duration. Measurements of each formant were taken at the eleven resulting points along its time axis.

Before any measurements could be made, however, segmentation rules were required to define what was meant by phoneme duration. For some of the spectrograms, segmentation of the utterances into their three component phonemes was obvious, but for others it was frequently ambiguous and subject to alternate interpretations. In addition, the specific goal of obtaining complete data on the vowel formants affected segmentation rules. For example, the necessity of having three clear vowel formants precludes segmenting plosive-vowel boundaries at the "spike," for the following release frequently obscures

the vowel formants. A second limitation upon segmentation is the sole use of broad-band spectrograms; other workers have frequently used supplementary information gained from narrow-band spectrograms, intensity curves, and any of a number of direct physiological measurements.

In general, except where precluded by data requirements or lack of displays beyond the broad-band spectrogram, rules for segmentation follow the recommendations made by PETERSON and LEHISTE [1960]. A detailed description of all the segmentation criteria along with a discussion of diverse phenomena encountered within the initial and final consonants appears in Appendix I. For convenience, a brief summary of the basic segmentation rules is presented here. As with all condensations, however, this summary is a simplification which neglects many complications actually dealt with in segmentation. The effects of segmentation rules on all the resulting durations discussed in this paper are extremely significant; references in the following sections are made to these rules on several occasions. Generally, though, the rules are the following:

1) Initial phonemes /#,f,θ,h,s,ʃ,p,t,k/ are segmented at the onset of voicing of the vowel; the same nine in final position are segmented at voicing termination.

2) Initial and final position /w,l,r,ʍ/ are segmented at the points of maximum rate of change of the second formant.

3) Phonemes /j,v,ð,z,ʒ/ in initial position are segmented at the point of strong intensity increase in the second formant; final position /j,z,ʒ/ are segmented at the point of second formant intensity decrease; final position /v,ð/ are segmented at the leveling out of the second and third formants from their decline.

4) Initial voiced plosives /b,d,g/ are segmented at the onset of the second and third formants; in final position, /b,d,g/ are segmented at the termination of these formants.

5) The three nasals /m,n,ŋ/ in both initial and final positions are segmented at the abrupt change in formant location or slope.

The actual process of measurement can now be considered. Hand measurements, utilizing dividers and a template, resulted in great accuracy in the hands of a skilled user, but quickly proved to be unacceptably slow. The three duration and 33 frequency measurements (11 frequency values for each of the first three formants) took an average of 20 minutes per spectrogram. As a result, the television system described below was developed and subsequently used.

A Digital Equipment Corporation PDP-8 computer is the central element in this spectrogram-measurement system. Attached to the computer is a Concord MTC-2 television videcon which is secured to a special mount which accepts spectrograms a controlled distance from the camera. A

computer then alternately displays the TV image of the spectrogram and an operator-directed system of lines and dots on a Fairchild 737A cathode ray tube screen. The two images succeed one another rapidly enough to merge them into one visually-integrated image. A light pen and teletype keyboard provide operator control of the system.

While developing and calibrating the system, it was observed that the TV camera is sensitive to room temperature, to nearby metal objects, and to what is apparently an internal temperature. This last factor was determined from the observation that for a constant room temperature, the system achieved a stable operation state after approximately 20 hours. Because of this, the camera was left on day and night for the entire measurement period. In addition to this, a continuing check on the system during operation was provided by measuring a new calibration value every three spectrograms. This procedure proved to be adequate, for the fluctuations in the system were considerably lower than anticipated.

Use of the computer-aided system for spectrogram measurement proceeds as follows. First, a drafted calibration grid derived from a typical 200 Hz tone spectrogram is placed before the TV camera and held against a steel plate by small magnets. Then it is manually adjusted so that the image of its bottom edge coincides with the lower of two horizontal lines generated by the computer. With the light

pen, the operator then moves the higher of the two lines until it coincides with the 3000 Hz line on the grid. The option to reset this calibration value before each spectrogram measurement, while available, was exercised only before each third measurement.

The calibration grid is next replaced by a sound spectrogram. This is likewise positioned so that its image base rests on the lower horizontal line. In addition, the leading edge of the vowel segment is aligned with the second of four vertical lines which also appear on the screen. The operator then segments the CVC triple by moving, with his light pen, the other three vertical lines into positions at the boundaries of the phoneme segments. When satisfied with the segmentation, the operator senses a dot with the pen, and the program shifts into the frequency-measurement phase.

The program divides the vowel segment into ten equal time segments and presents the operator with three horizontal rows of 11 points each. These points represent the ten segments of each of the three formants. With the light pen the operator adjusts the points vertically until they form an acceptable representation of the vowel formants' images. He then senses another dot and types the adjustment factor previously determined for that spectrogram. The adjustment factor along with the duration and frequency measurements

are automatically recorded and printed out for verification. The operator then proceeds to the next spectrogram.

Utilization of this television system has resulted in a slight decrease in accuracy over hand techniques, but has accelerated processing time by a factor of four. The 20-minute period required for hand measurement has been reduced by the TV system to five minutes per spectrogram. Additionally, the data, consisting of 5184 durations and 57,024 formant frequency values, are already recorded by the computer.

An analysis of the duration data is presented in the remainder of this paper.

Duration Distributions

An examination of the duration distributions can provide not only a useful overview of the data, but also information on specific phonemic qualities. Figure 1 contains plots of the basic distribution of each of the three phonemes--the vowel and the two consonants. In order to avoid all the zero-valued "durations" that the condition of silence yields, as well as to separate the phoneme pairs from the triples, only the 1587 durations for each phoneme from full triples are shown here.

The distribution of C_2 is quite different from that of both C_1 and V. Not only are its durations substantially

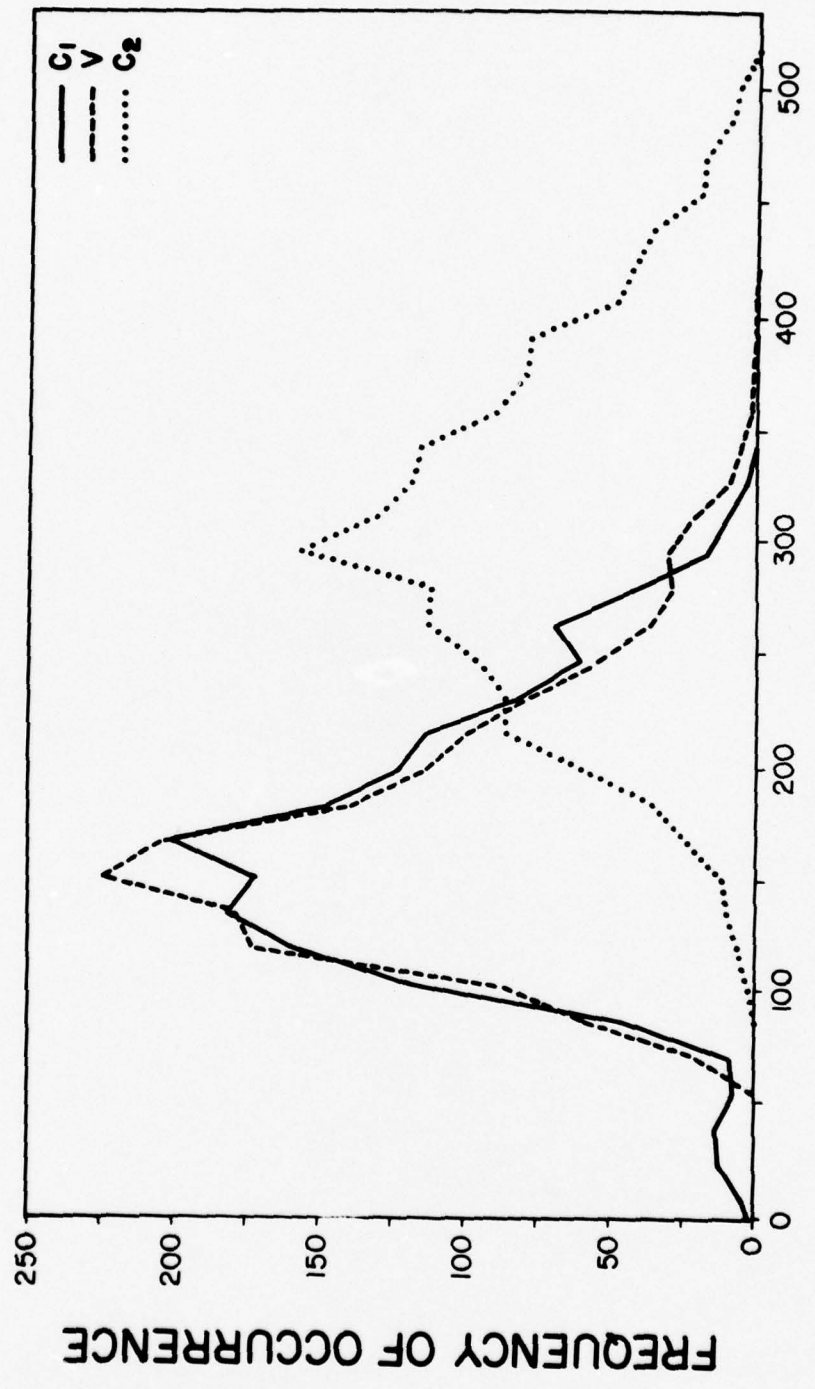


Figure 1. Duration distributions for the initial consonant, the vowel, and the final consonant pooled for all 1587 items not bounded by silence.

longer than those of C_1 and V, but the curve is also more symmetrical and less skewed toward the lower durations. In the C_1 distribution, the drops in frequency of occurrence around 150 msec and 250 msec, as well as the cluster of extremely short durations below 50 msec, are all explicable in terms of the set's internal composition, as shall be seen below.

The approximation of these three curves and their three component days to normal distributions were determined by the Chi-Squared Goodness of Fit Test. The two principal observations emerging from this are: 1) distributions of C_1 and V are not normal, while C_2 is normal; and 2) all three of the individual component days for C_1 and V are more nearly normal than their composite distributions, while, for C_2, only one of the three days is more nearly normal. In order to better understand the distributions of each phoneme, the more detailed analyses which follow were carried out.

Initial Consonant Distributions

The distribution of the first consonant, separated into its three component days, is presented in figure 2. Several characteristics are immediately apparent: 1) short durations below 50 msec appear in all three day groups; 2) data for the second day are much more skewed toward short durations

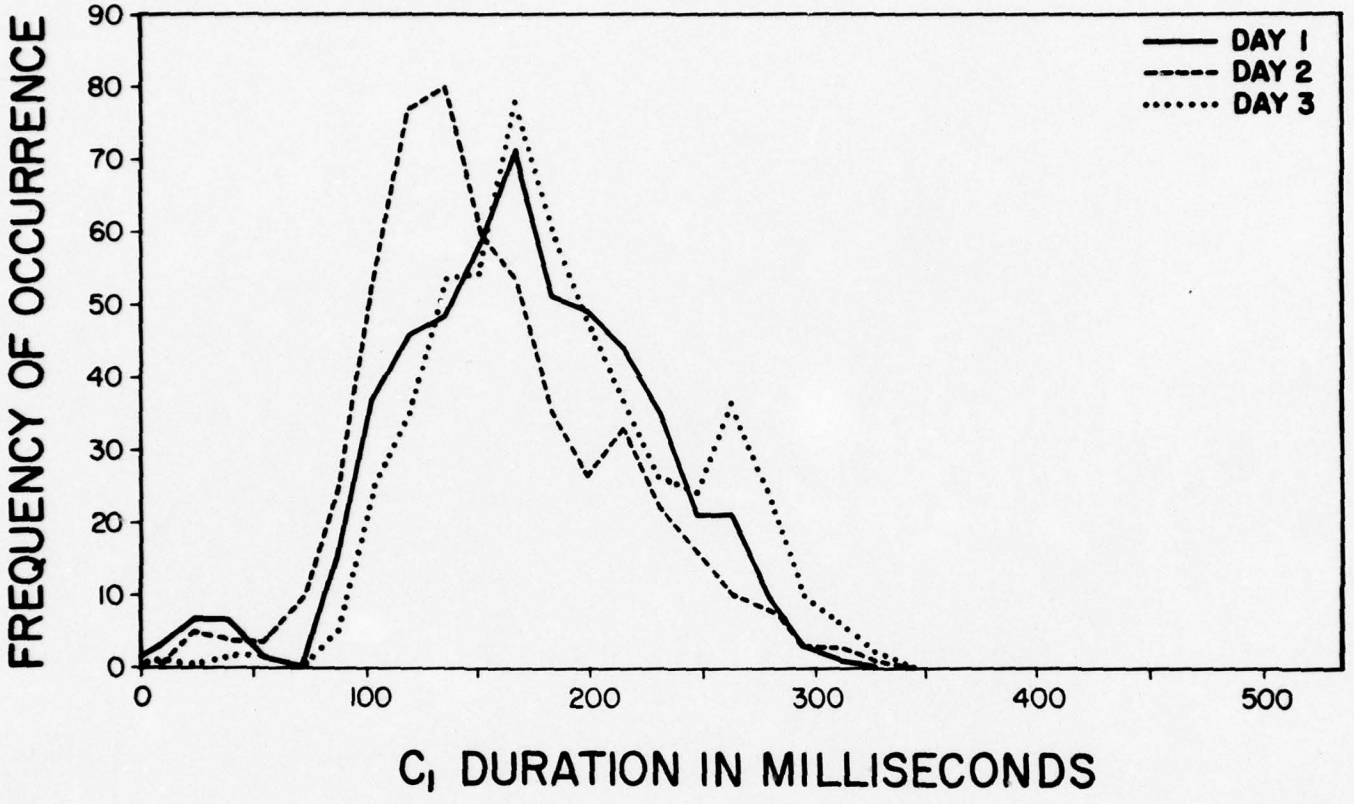


Figure 2. Duration distributions for the initial consonant measured from the 3 recording days.

than data for the others; and 3) all three days have a sharp dip located between 200 and 265 msec. These three observations can be seen to be related to the three dips observed in the C_1 distribution of figure 1. Two of the three dips--the high and low-- apparently occur in all three days, while the dip appearing in the middle of the composite C_1 distribution is attributable to the mismatching of the three individual day peaks.

Figure 3 represents the results of separating the C_1 distribution into two groups, depending upon whether C_1 is a voiced or voiceless consonant. Here the nine voiceless consonants are plotted below the central line, while the 14 voiced ones lie above. Several observations may be readily made: 1) the low-duration occurrences are principally, although not exclusively, voiced; 2) the voiceless consonants are generally longer in duration than the voiced ones; 3) the voiced consonants are fairly normally distributed, while the voiceless ones are heavily skewed toward the right; and 4) the long-duration dip occurring for all three days in figure 2 is the apparent result of an early decrease in the voiced distribution concurrent with a still-rising voiceless group.

Similarly, the C_1 distribution can be separated into the consonants' five manners of articulation, as plotted in figure 4. It is apparent from these three individual day plots that the distributions have a fairly high

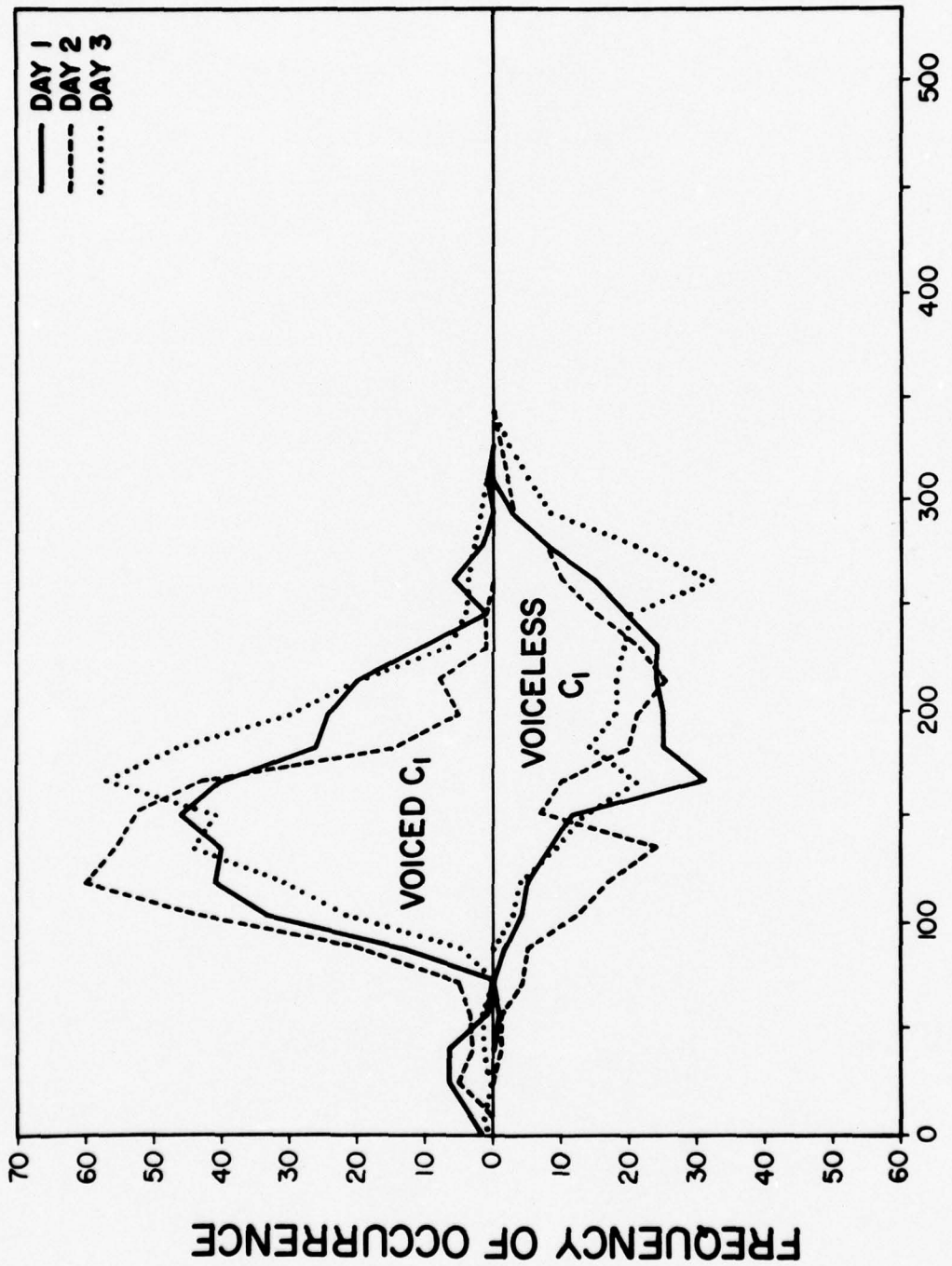


Figure 3. Duration distributions for the initial consonant measured from the 3 recording days and separated according to the consonant voicing.

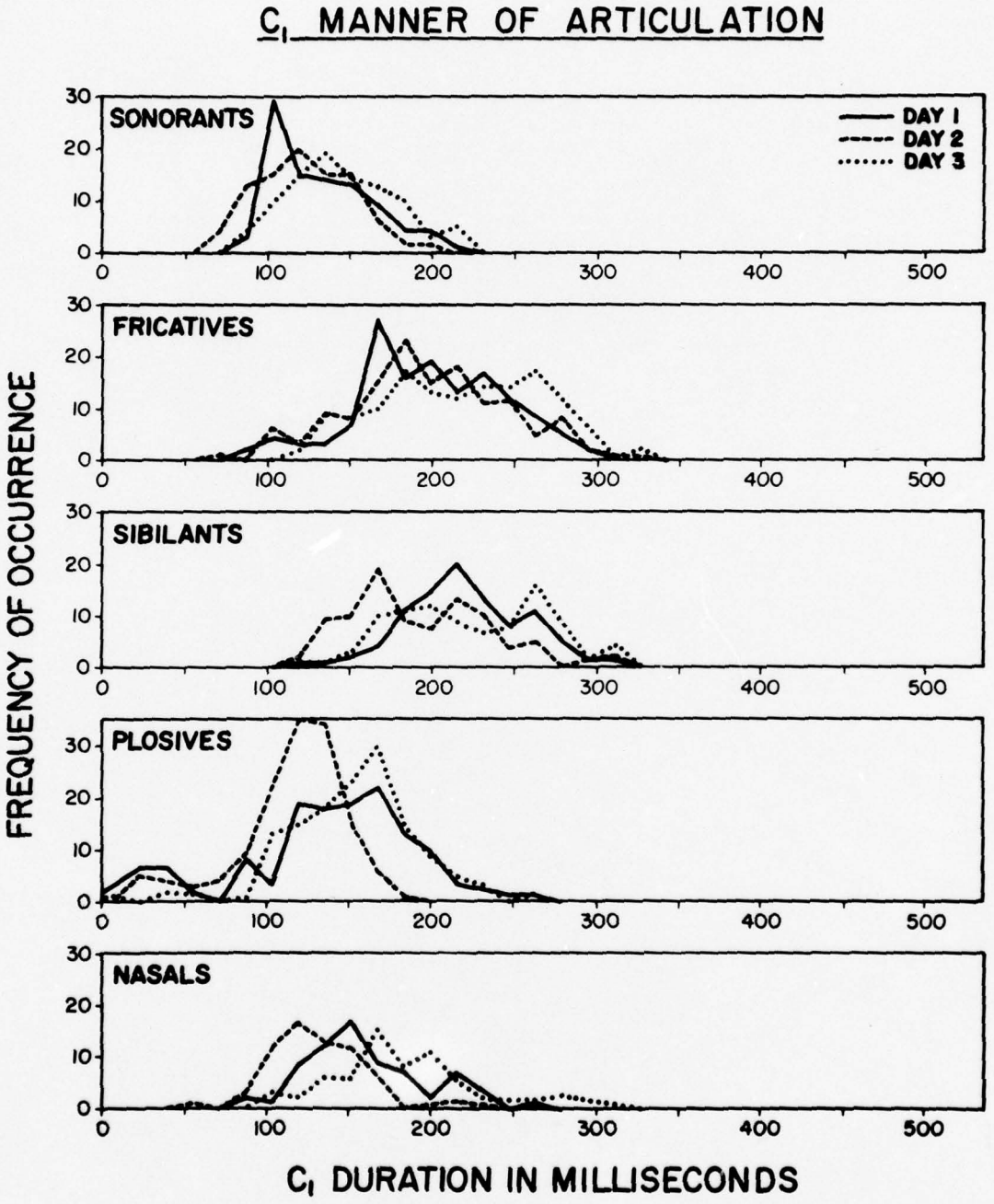


Figure 4. Duration distributions for the initial consonant organized according to the consonantal manner of articulation. Data from the 3 recording days are plotted separately.

repeatability. Over all five manners, the second day durations are the shortest, and those from the third day, the longest. Distributions for the first day are similar to those of the second day for the sonorants and fricatives, but are more similar to those of the third day for the plosives, while in the sibilant and nasal categories the first-day distributions are different from those of both other days. The reason for such variation is not known. The observation, however, that for all five manners of articulation, the durations from the second day are consistently the shortest implies that the utterances on that day were spoken more rapidly than on the other days. This situation, however, fails to hold for the vowel and final consonant, as will be seen later in this section.

The plosives as they appear in figure 4 are obviously the only candidates for the very short C_1 durations which have been appearing. A closer examination of the individual plosives is presented in figure 5; here again, the voiced phonemes are above the voiceless ones. Out of the 38 plosives having durations of 70 msec or less, 20 are /g/, 12 are /d/, 4 are /p/, and 2 are /b/, i.e., all but 4 are voiced. The observation that the voiced plosives are dichotomized into two separate duration groups is in accord with the separation found in voiced plosive duration data by LISKER and ABRAMSON [1967].

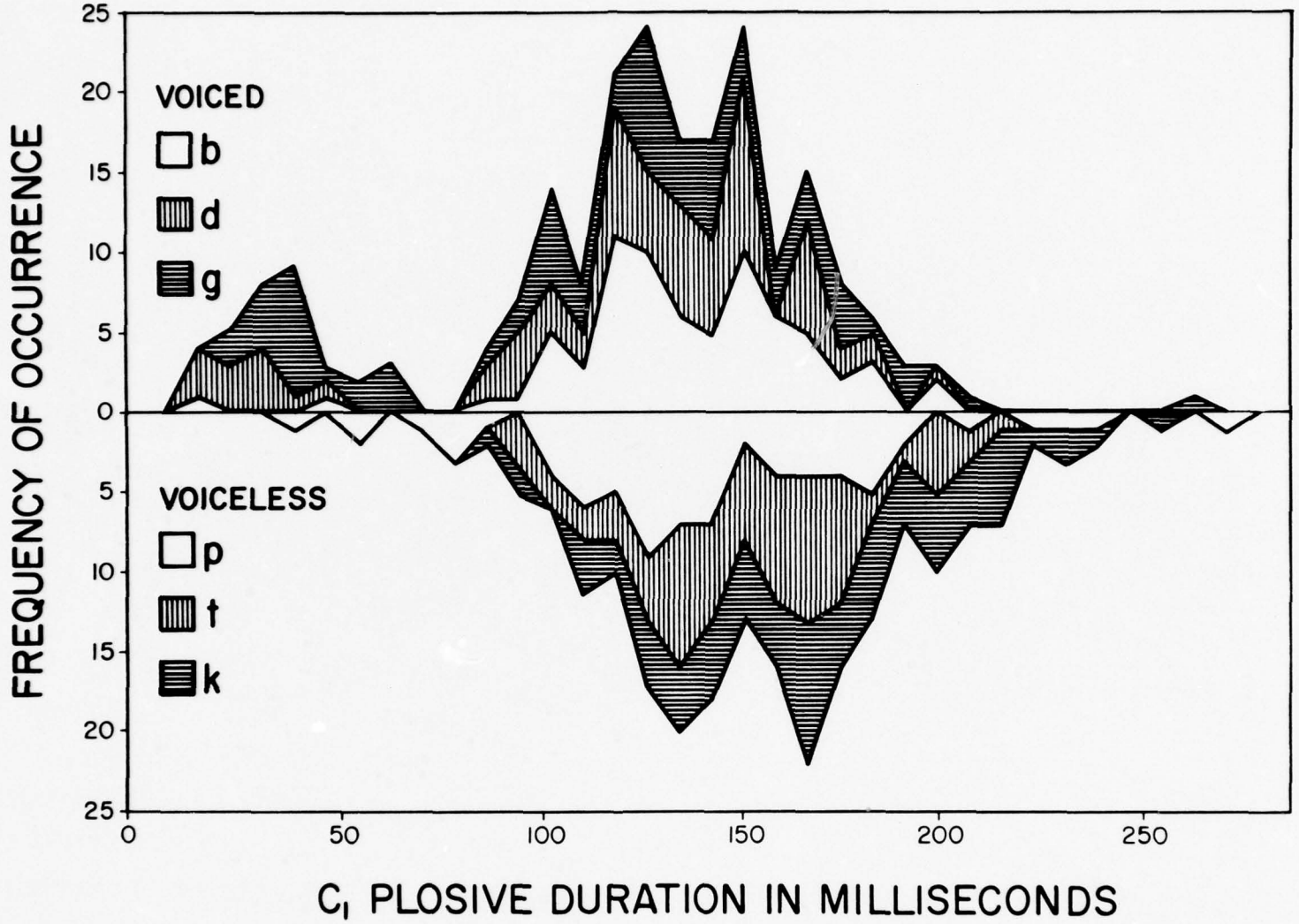


Figure 5. Duration distributions for initial plosives. The voiced plosives are plotted above the centerline, the voiceless ones below. The distribution for each plosive is plotted separately. Data from all 3 recording days are pooled

The fact that some plosive phonemes are considerably shorter than others is apparently attributable to a large extent to that phoneme's voicing characteristic and possibly to the articulatory position--the back plosives are generally shorter than the front ones. A question arises: does C_2 identity also influence the C_1 plosive durations?

An analysis of the 38 CVC triples where $C_1 \leq 70$ msec reveals the following: 1) A voiced C_2 is 1.18 times as likely to be associated with a short C_1 plosive as a voiceless C_2 is. 2) The frictional phonemes /ʍ f v θ ð h s z ʃ ʒ/ are 1.66 times as likely to be associated with short C_1 plosives as the more laminar ones /w l r j m n ŋ/ are. The C_2 plosives fall between the other two frequencies. And, most interesting, 3) the C_2 phonemes articulated in the back of the vocal tract are more highly associated with short C_1 plosives than C_2 phonemes articulated in the front. This third observation is graphically displayed in figure 6. There, the number of occurrences of $C_1 \leq 70$ msec per C_2 phoneme is plotted against C_2 horizontal place of articulation.

In general, it seems reasonable to conclude that shorter C_1 plosives are determined primarily by C_1 voicing state and C_1 place of articulation, partially by C_2 place of articulation and C_2 breath stream pattern, and possibly also by C_2 voicing state.

The existence of these short plosive durations is nevertheless a difficulty through much of this study. For

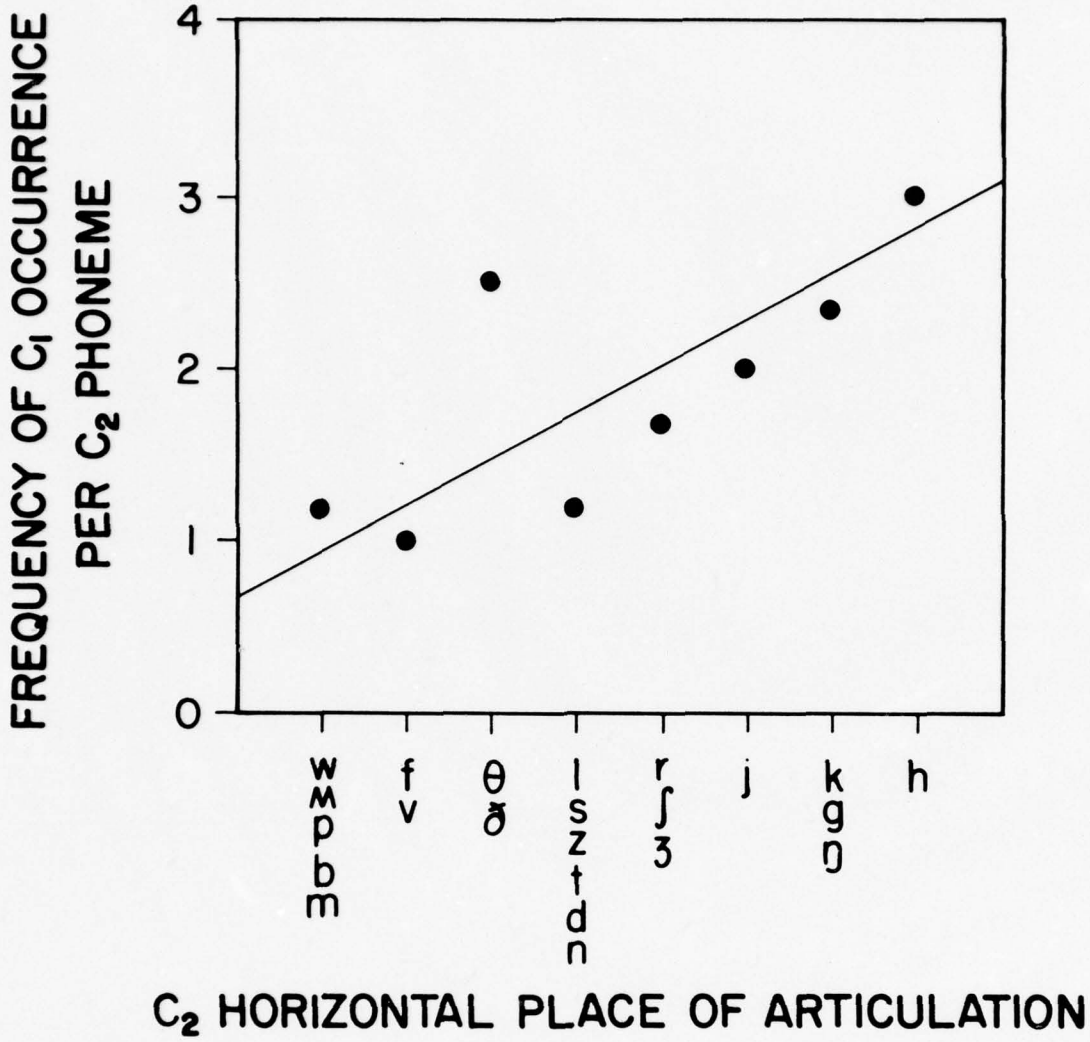


Figure 6. Frequency of occurrence of short (duration ≤ 70 msec) initial consonants plotted against the horizontal place of articulation of the associated final consonant.

all other speech sounds examined in this research, there is a high correspondence between physiological movement and audible speech. In the initial plosives, however, the articulatory action and buildup of pressure commence appreciably in advance of any audible result. Physiological measurements were not made, so the periods of plosive closure are unknown for C_1. Further, since C_2 is always preceded by a vowel, the closure period is always included in C_2 plosive durations, making C_1-to-C_2 comparisons difficult.

Vowel Distributions

Consider the vowel distributions of the three individual day sets as they appear in figure 7. From one day to the following, there is a plain progression from shorter toward longer durations. This contrasts with the day-to-day shifts within the first consonant, but tends more toward agreement with the C_2 day variations. The basic differences are summarized in table I. These values are derived from the 529 CVC triples per day not containing the element of silence /#/. For averages over so many measured durations, the lack of accord among the three days is both striking and perplexing. Since there appears to be a greater correspondence between V and C_2 than between C_1 and either V or C_2, one tends to look critically for an unbalancing factor within the C_1 set. The day-to-day C_1 duration positions, however,

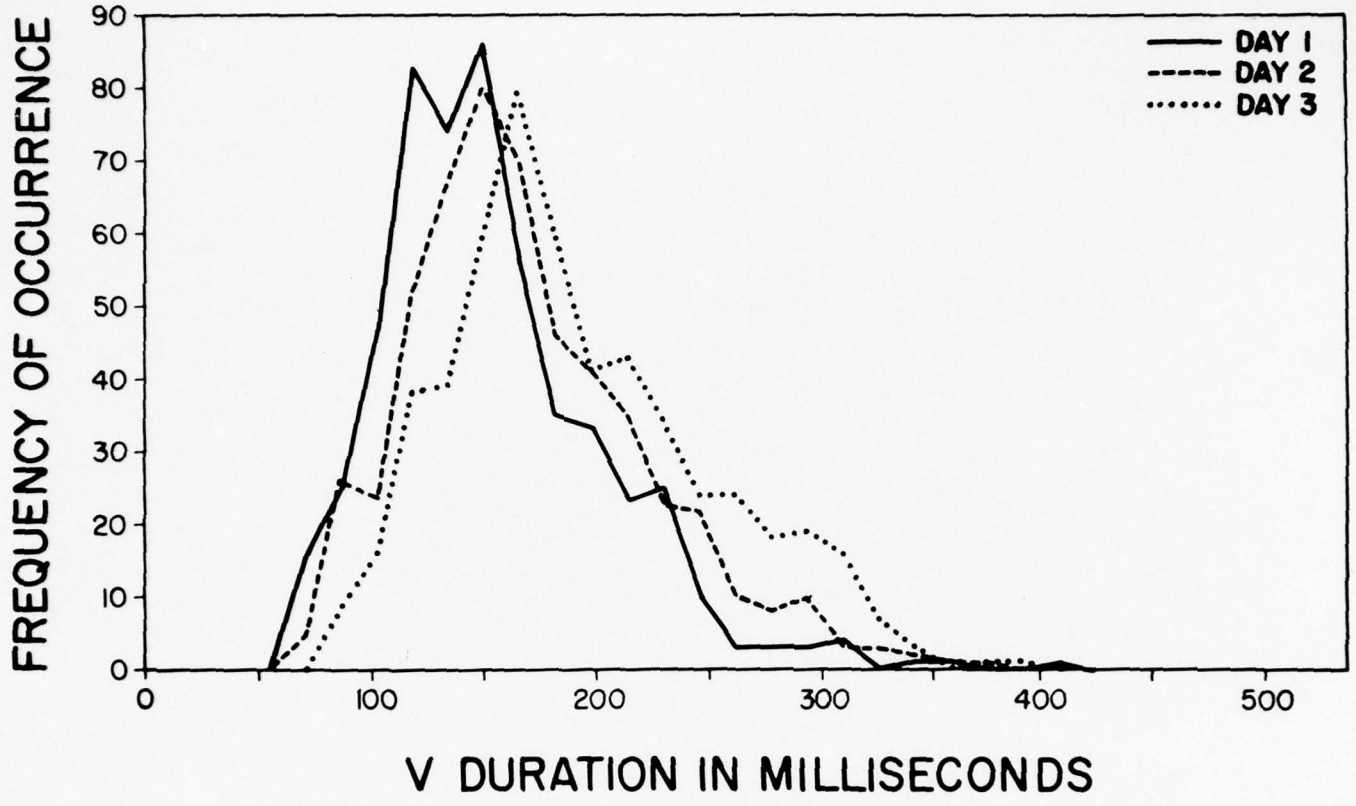


Figure 7. Duration distributions for the vowel measured from the 3 recording days.

	MEAN DURATION		
	C_1	V	C_2
DAY 1	169.4 ms	153.6 ms	294.7 ms
DAY 2	154.2	170.5	309.8
DAY 3	185.2	193.2	311.4

Table I. The average segment duration for each recording day.

hold intact over both voicing and manner of articulation analyses as has just been seen. The possibility of phoneme duration ratios assuming varying values on different days depending upon the mood of the speaker should be considered, but pursuit of this is outside the scope of this study.

Unlike the consonants, the vowel /ɪ/ has only one state of voicing and cannot be separated into voiced and voiceless components as was done for C_1. It can, however, be split into groups of durations depending upon the voicing characteristic of either C_1 or C_2, yielding information on the influence of adjacent consonants. In figures 8 and 9, distribution plots of the three recording days are presented according to the voicing characteristics of C_1 and C_2 respectively. In general, the following may be observed: 1) in both cases, the voiceless consonants are associated with shorter vowels, and the voiced consonants, longer ones; 2) this short-long voiceless-voiced contrast is the opposite of that which the C_1 voicing characteristic has upon its own duration; 3) the agreement of the distributions from one day to another is more consistent here than for C_1, this is especially true for the C_2 influence; 4) the effect of C_2 upon the vowel is greater than that of C_1, producing a greater differentiation between vowel durations under voiced and voiceless C_2 influence; 5) all plots are skewed toward lower values, but those under voiceless

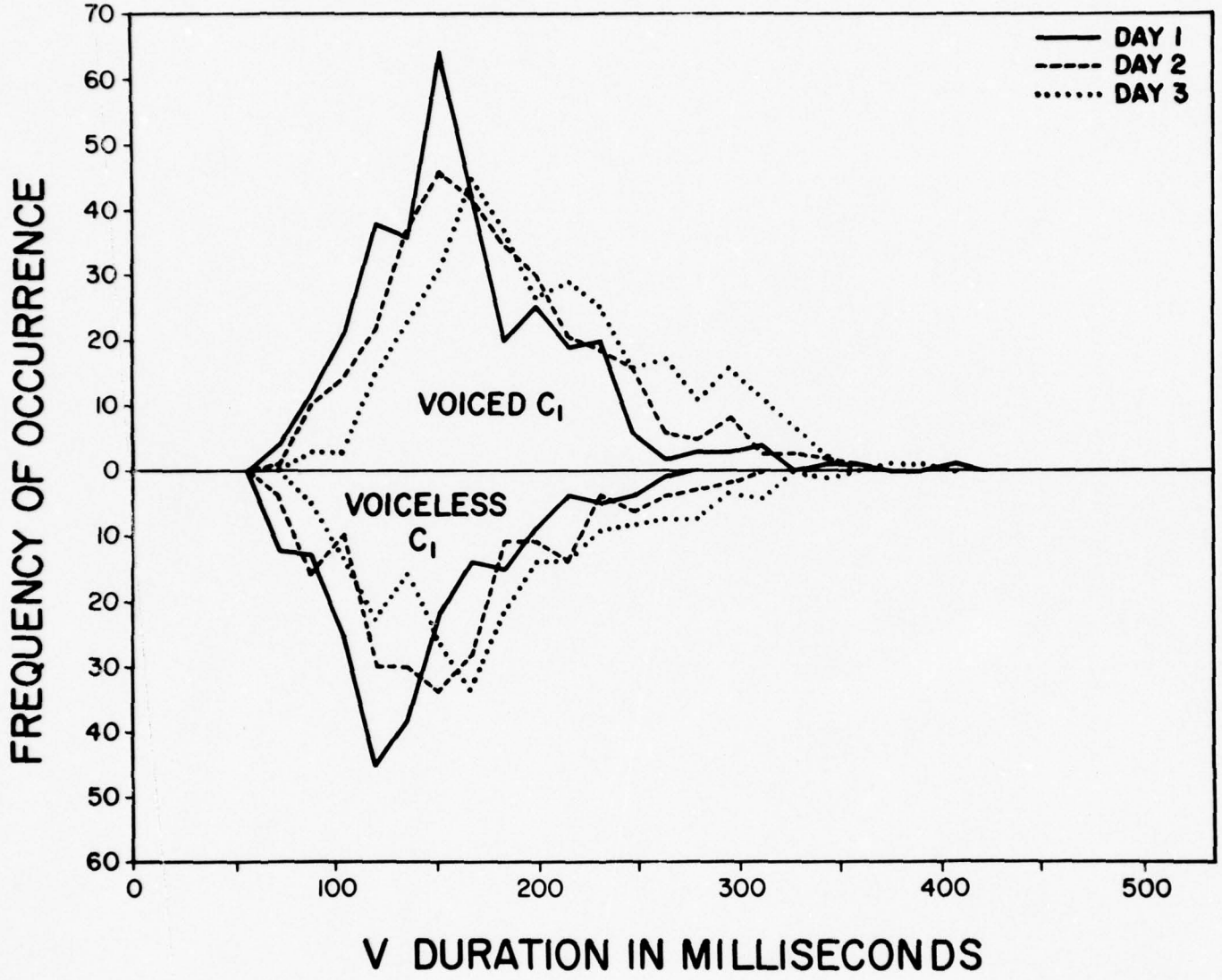


Figure 8. Duration distributions for the vowel plotted according to the voicing of the initial consonant. Data from the 3 recording days are shown separately.

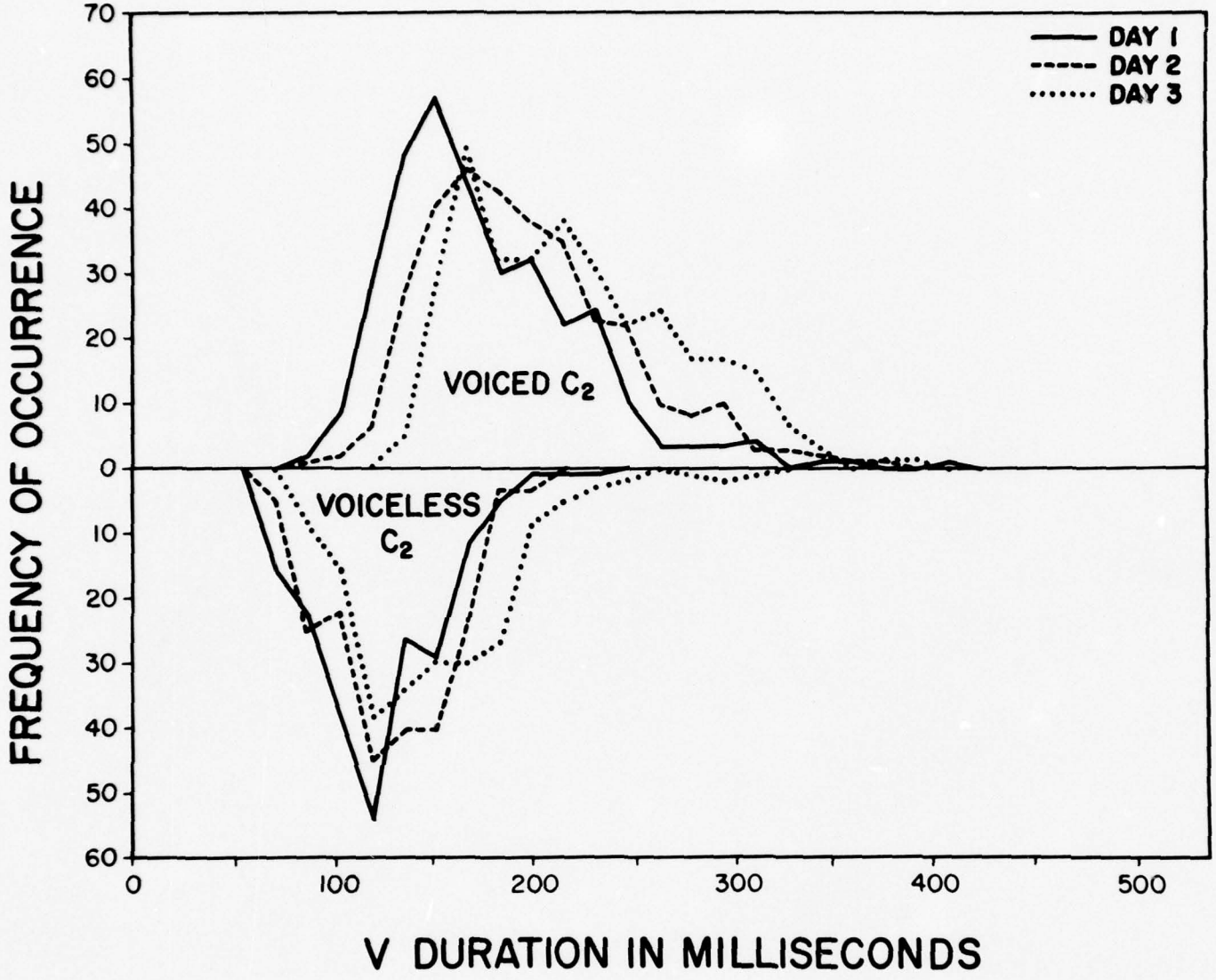


Figure 9. Duration distributions for the vowel plotted according to the voicing of the final consonant. Data from the 3 recording days are shown separately.

consonantal influence appear to have more normal distributions; and 6) the voiceless consonants' influence upon the vowel results in a very closely bunched duration group, but the voiceless C_1 distribution itself has an extensive range, apparently with two separate peaks. A re-examination of the total vowel distributions of figure 7 reveals that (the total) plots are composed of phonemes associated with voiceless consonants in the low-duration range and those associated with voiced consonants in the high-duration range. The fact that the distributions in all three vowel figures have a short-duration skew shows that the skewing occurs at a more fundamental level than is presented here; e.g., it is not the result of pooling the voiced and voiceless consonant sets, for the vowels also manifest it.

The vowel durations may also be separated according to manner of articulation of the adjacent consonants. Figures 10 and 11 are distribution plots of the vowel durations on the three days, depending upon the articulation manners of C_1 and C_2, respectively. It may be observed from the figures that, as with the voicing differentiation, manner of articulation of C_2 has a greater influence upon vowel duration variation than does C_1; the plots in figure 10 appear more similar to their parent plots of figure 7 than do the plots of figure 11. One feature in figure 11 is the clustering effect of C_2 nasals upon vowel durations. They are more closely bunched than any of the other plots

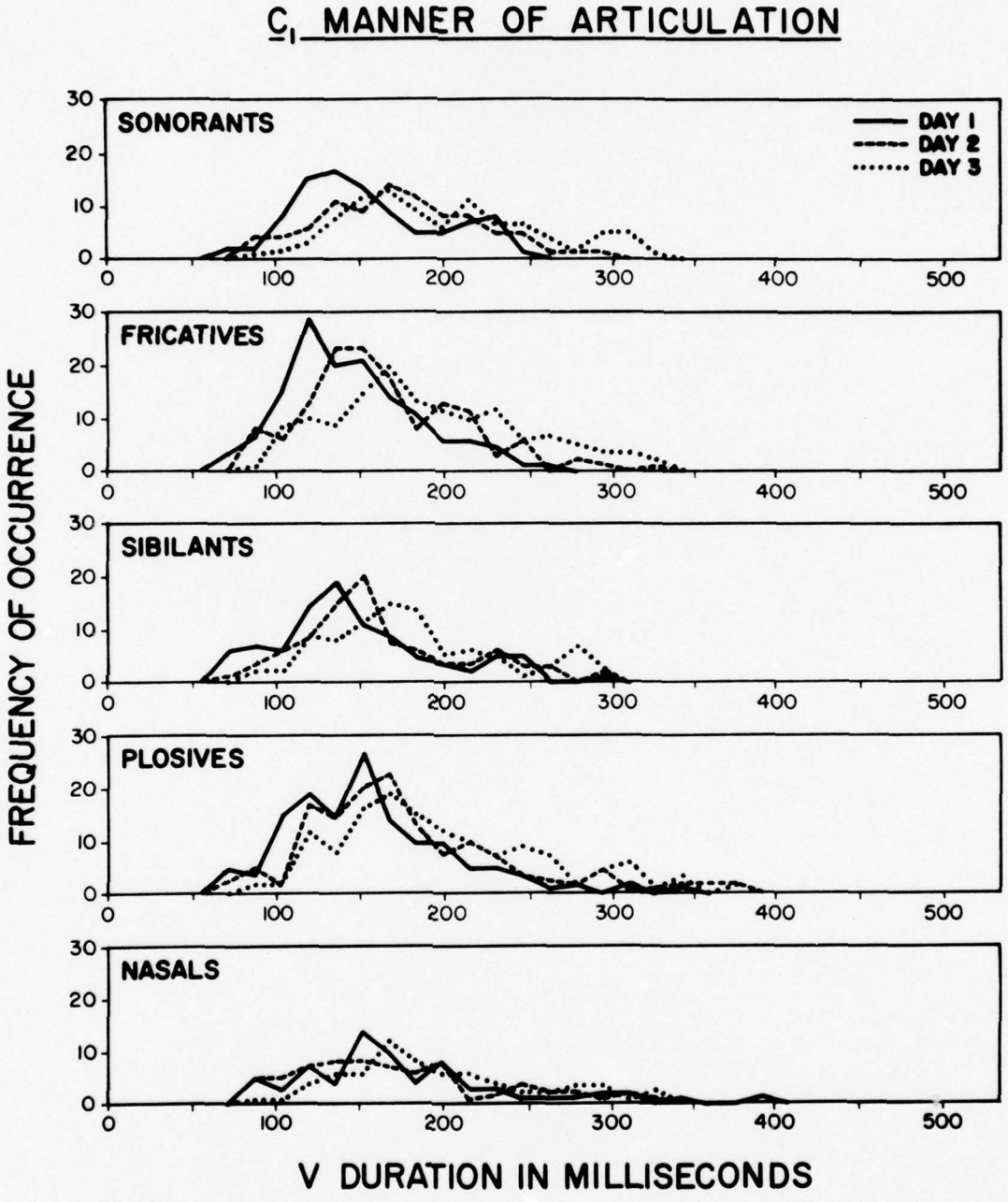


Figure 10. Duration distributions for the vowel organized according to the manner of articulation of the initial consonant. Data from the 3 recording days are shown separately.

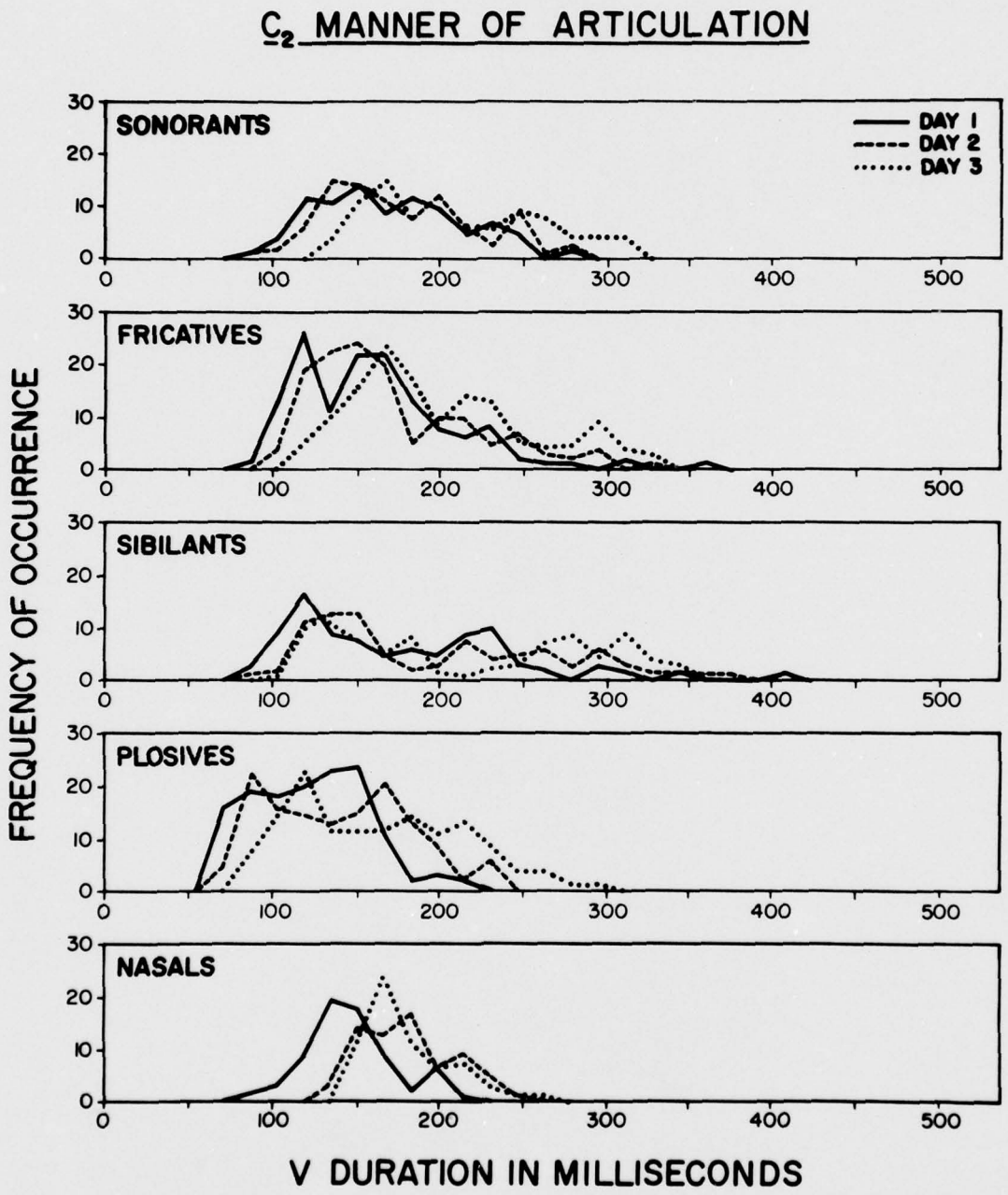


Figure 11. Duration distributions for the vowel organized according to the manner of articulation of the final consonant. Data from the 3 recording days are shown separately.

in the two figures. By contrast, C_1 nasal influence upon the vowel duration produces the flattest, most spread-out distributions of all. Conceivably the strong control exerted by C_2 nasals upon vowel durations serves as a cue to a listener that the subsequent consonant is a nasal; also the clustering could perhaps be the result of physiological necessity.

Final Consonant Distributions

The second consonant distribution, characterized by long durations and a near normal distribution, is split into its three component days in figure 12. Although they have many local maxima, the three individual days can be seen to follow one another closely, each having an approximately normal configuration.

The division of these C_2 durations into voiced and voiceless categories results in the distributions which appear in figure 13. The following observations may be made: 1) the highly normal distributions for C_2 in general consist of two sets, voiced and voiceless, neither of which is as nearly normal; and 2) the skewing on the two groups is similar to that on C_1 and dissimilar to that on V under the influence of consonantal voicing.

A further analysis of C_2 durations is made in figure 14 where the consonants are separated into five manners of

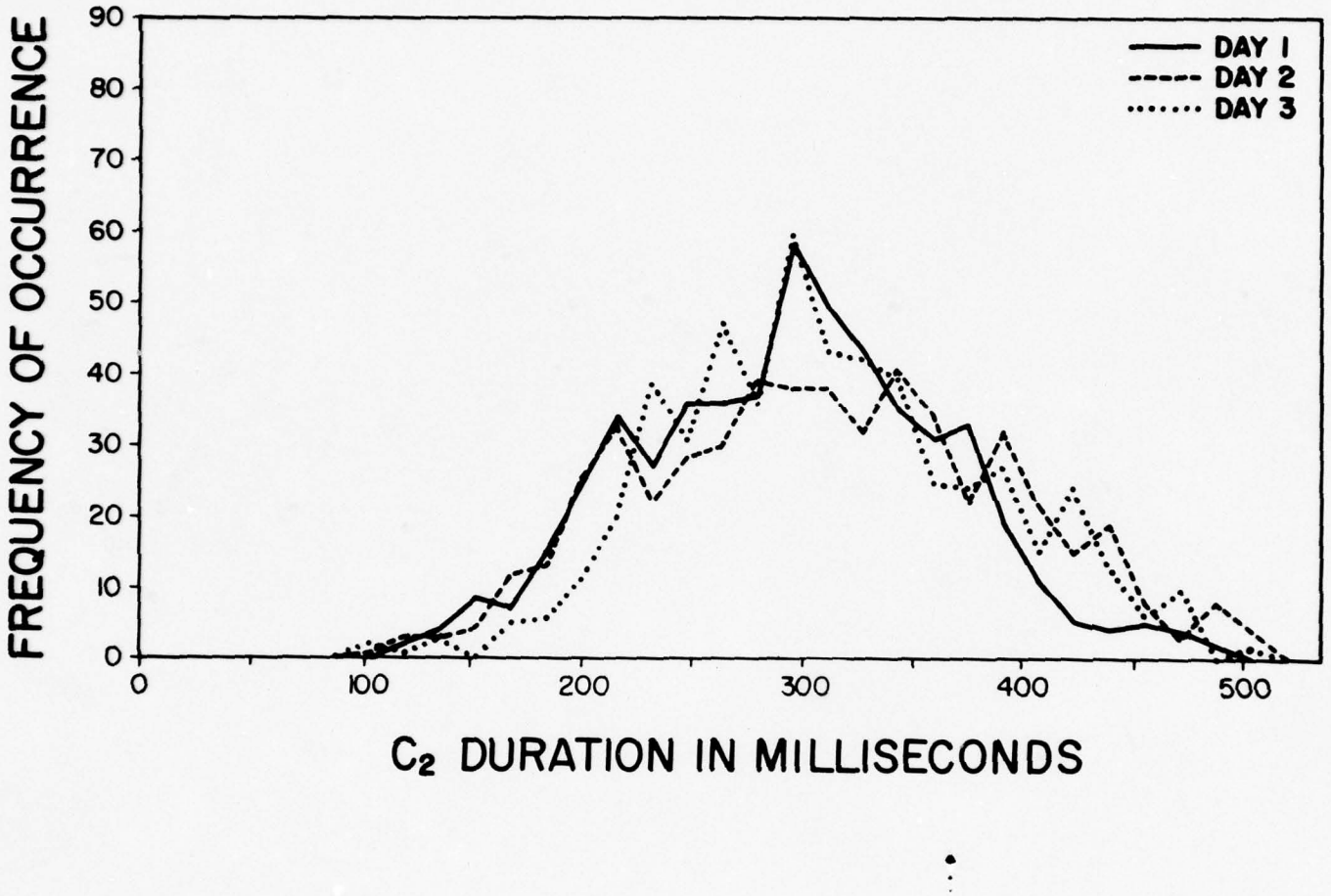


Figure 12. Duration distributions for the final consonant measured from the 3 recording days.

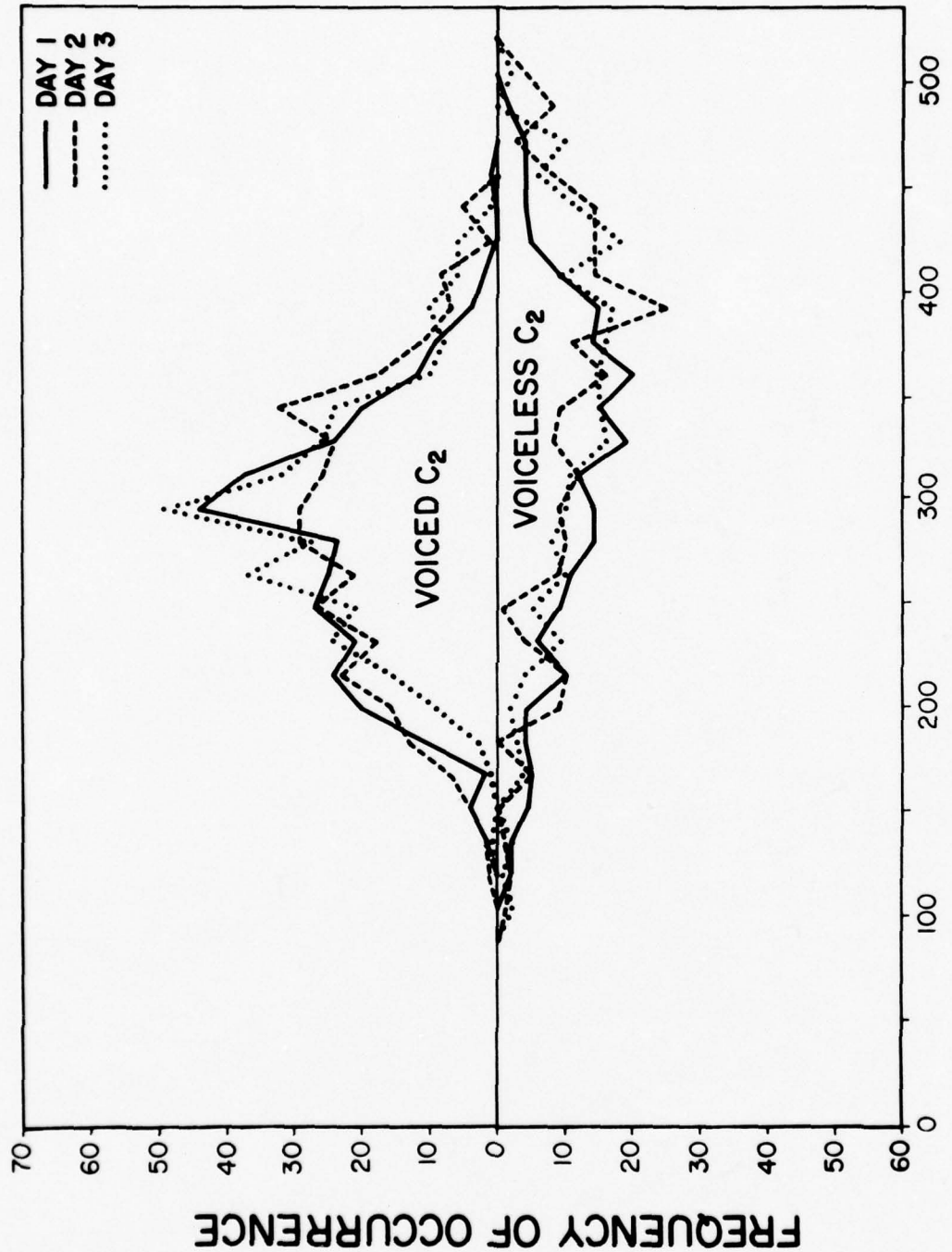


Figure 13. Duration distributions for the final consonant plotted according to the consonantal voicing. Data from the 3 recording days are shown separately.

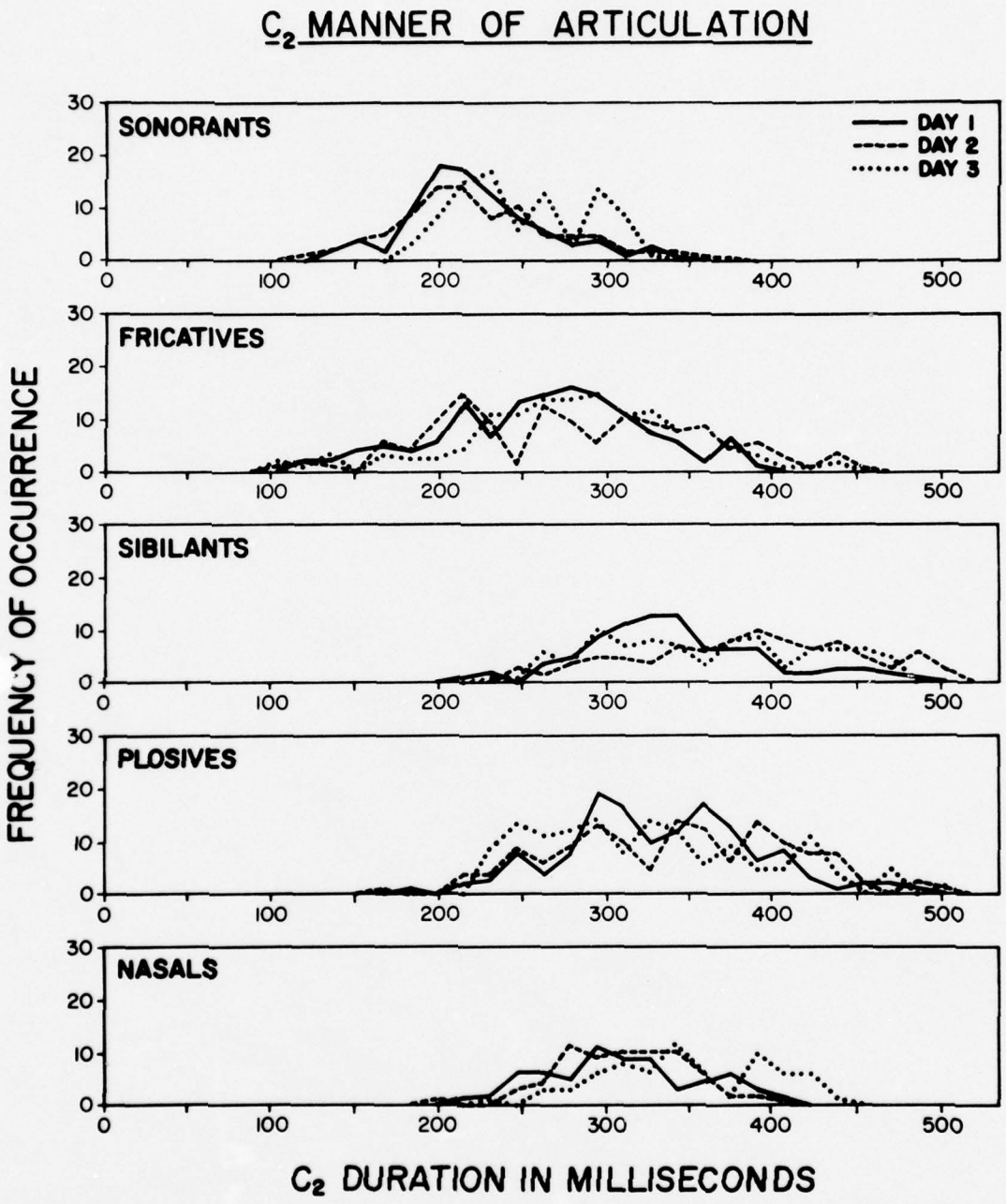


Figure 14. Duration distributions for the final consonant organized according to the consonantal manner of articulation. Data from the 3 recording days are shown separately.

articulation. As in the case with C_1 durations, there is a distinct clustering around certain durations for each of the manners of articulation. The very low plosives in C_1 contrast sharply with the C_2 distributions, not only in that the C_1 durations contain an extremely low group, but also in that they are highly concentrated in general around low durations, while the C_2 plosives tend to have a wide range. The locations of individual manners of articulation relative to their respective C_1 or C_2 ranges vary somewhat. The nasals are shorter than the fricatives in C_1, but longer in C_2, for example.

In general, the following observations on phoneme distribution are most important:

1) Phoneme distributions for C_1 and V are highly similar and strongly skewed toward low durations.

2) C_2 has significantly longer durations and an approximately normal distribution.

3) Voiced consonants tend to have shorter durations than voiceless consonants.

4) The effect of voiced consonants on their adjacent vowels is to lengthen them, while voiceless consonants shorten them.

5) Manners of articulation have certain different duration ranges for the two consonant positions.

6) The mean C_1:V:C_2 duration ratios vary significantly from one day to another.

Analysis of Variance

Although the analysis of distributions provides general qualitative information on the effects of consonantal identity upon the durations of the consonants and the intervening vowel, it cannot provide answers to several questions. For example, does C_2 identity influence C_1 durations, and vice versa? How statistically significant are the duration-altering influences of the consonants upon vowel durations, upon their own durations, and, if they exist, upon the other consonant durations? Can a mathematical model be constructed to describe and predict the effects of consonantal phoneme identity upon durations? How accurate can such predictions be? These and other questions are answered in this section by an application of analysis of variance and some associated statistical techniques to the data.

The purpose of this section is to construct and test a mathematical model which describes segmental duration as an overall mean segment duration which is lengthened or shortened by the effects of the initial and final consonants. The hypotheses tested are: 1) there is no effect on segment duration due to the initial consonant; 2) there is no effect due to the final consonant; 3) there is no effect due to an interaction between the two consonants. Which of these hypotheses are accepted or rejected will then be used to

determine which components of the model will be retained. The statistical technique used is a two-way fixed-effects model analysis of variance.

Some mathematical notation is required to facilitate identification of the durations. First, let the 23 consonants and silence /#,w, l, r, j, ɱ, f, v, θ, ð, h, s, z, ʃ, ʒ, p, b, t, d, k, g, m, n, ŋ/ be numbered 1 through 24, respectively. Now define phoneme duration $D^{p}_{ij/k}$ whereby

$p = C_1, V_1, C_2$; one of the three phonemes

$i = 1,2,\cdots,24$; the number of C_1 identity

$j = 1,2,\cdots,24$; the number of C_2 identity

$k = 1,2,3$; the recording day number.

As an example, $D^{C_2}_{2\ 24/3}$ symbolizes the duration of the second consonant of the third recording of /wɪŋ/.

It was decided to delete from consideration all "triples" which contain the phoneme of silence /#/ in any position. This was done not only to create a more homogeneous body of data, but also to avoid problems in associating zero-variance durations with the assumptions basic to analysis of variance.

The model which is used is:

$$D^{p}_{ij/k} = \mu^{p} + a^{p}_{i} + b^{p}_{j} + c^{p}_{ij} + e^{p}_{ij/k}$$

where μ^p is an overall mean, a_i^p is an effect due to C_1, b_j^p an effect due to C_2, c_{ij}^p is an interaction effect, and $e_{ij/k}^p$ is an error term, which is assumed to be normally distributed with zero mean. The interaction term is an effect due to the C_1 and C_2 effects acting other than strictly additively.

For each phoneme, C_1, /ɪ/, and C_2, the data fill a 23 x 23 matrix, on one axis are the identities of the C_1 consonant; on the other, of consonant C_2. In each of the 529 cells are three individual duration observations, one from each of the three recording days.

It was decided in this situation to utilize a two-way fixed-effects model (rather than a random-effects model), entailing tests for interaction. The principal differences between "fixed-effects" and "random-effects" models lie in their fundamental assumptions rather than in their techniques. The latter model entails the assumption that the row and column categories are selected at random from a larger population, and the conclusions obtained from the analysis pertain to the population as a whole. On the other hand, the fixed-effects model does not consider the possible population of categories from which those being analyzed were taken, but only the categories themselves; the inferences drawn from the analysis, then, apply only to that set of categories. Since the categories in this study consist of the consonantal phonemes of English,

and since these categories are not random samples from a large population (say, of possible human consonantal sounds), a fixed-effects model is more appropriate.

Since the data were recorded in triplicate, a test for an interaction effect between C_1 and C_2 is possible. From a theoretical point of view, it is certainly conceivable that under certain circumstances, interaction could occur.

Before entering the analysis of variance itself, it is desirable to look at the associated assumptions. According to BLALOCK [1960], they are the following:

(1) Samples are random and independent.

(2) Cell, row, and column populations are normal.

(3) Subcell variances are equal.

Assumption (1) was not tested in any specific way, but deemed to be generally satisfied upon the observation that the speaker on three different days recorded the utterances in three different orders.

The part of assumption (2) involving normality of the cell populations was likewise not tested because each cell contained only three observations; this is too few to check. The rows and columns, however, contained 69 durations each and therefore readily lent themselves to chi-square comparison with normal distributions. Accordingly, the 138 row and column distributions in the study were compiled and tested by computer programs. Results show that 121 out of the 138 distributions are normal at the .05 level. Generally speaking,

the very large number of normal row and column distributions is sufficient to accept assumption (2).

As for the homogeneity of subcell variance entailed in assumption (3), it was found that extensive variation exists within the variances. Values range from 1.6 to 11836.2, although the high variance within C_2 cells is considerably higher than that in either C_1 or V groups. Figure 15 illustrates the distributions of variances for the three phoneme durations. It may be observed that vowel variances are essentially lower than C_1 variances, which in turn are substantially lower than those of C_2. Since the C_2 durations are longer than C_1 durations, the larger variances are not surprising. What may be unexpected is the realization that vowel variances are not much lower than C_1 variances, especially when one considers that the vowel is always /ɪ/, whereas C_1 ranges over 23 possible consonants.

In order to check the homogeneity of variances within each of the three groups, Bartlett's test [DIXON and MASSEY, 1951] was conducted. Results showed that group C_2 is the most nearly homogeneous of the three. There the chi-square test revealed that C_2 variances are just outside the .05 level, which is the usual comparison level employed by Bartlett's test, but it is nevertheless reasonable. As for the groups of C_1 and V variances, Bartlett's test found them substantially outside the .05 level. Obviously, assumption (3) cannot be met by these data. This does

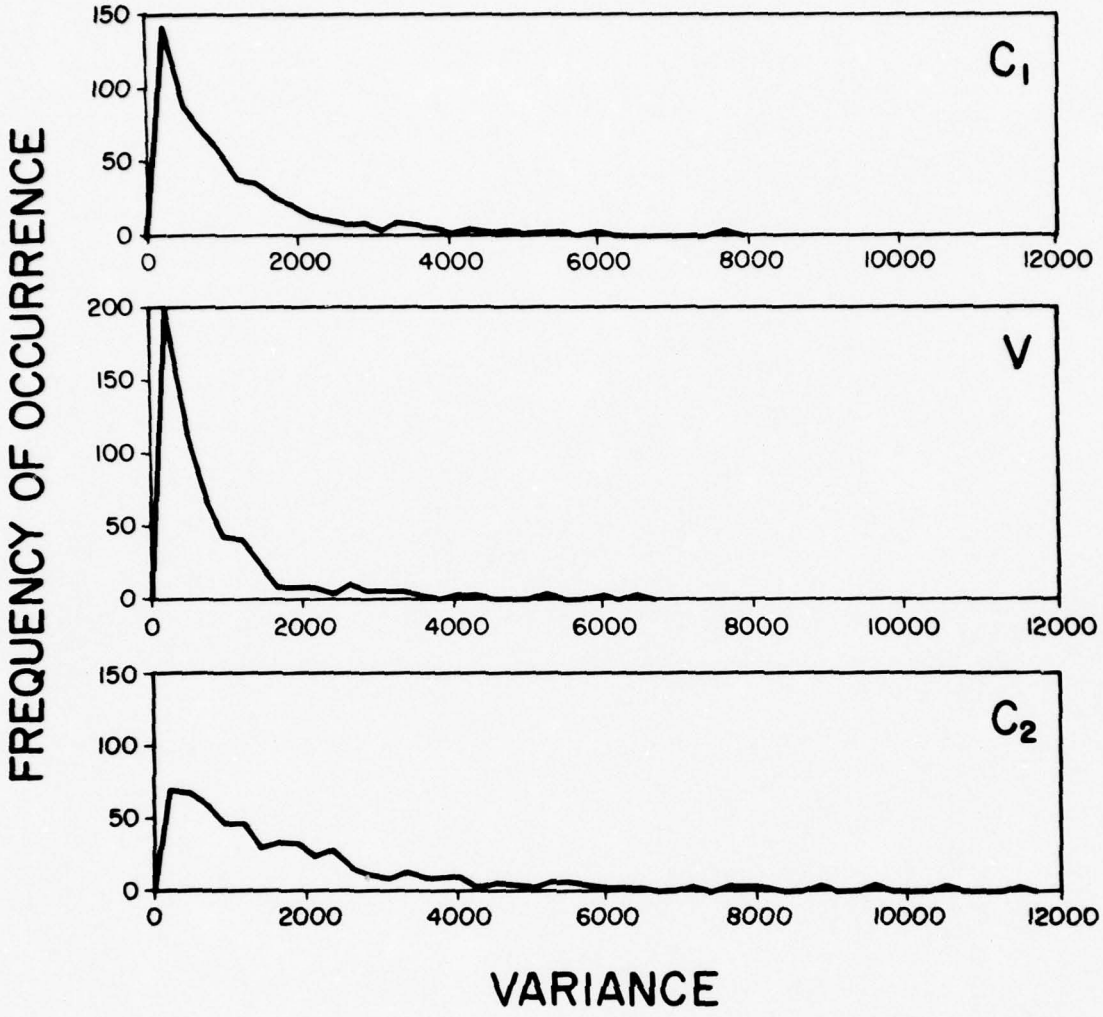


Figure 15. Distributions of the variances of the 138 row and column distributions for the initial consonant, the vowel, and the final consonant.

not, however, result in abandonment of analysis of variance in this circumstance, because departure from the basic assumption (3) can be sustained without radically altering the resulting inferences. This is particularly true when conclusions are limited to inferences about fixed main effects and interactions, and when, as in this case, cell numbers are equal [SCHEFFÉ, 1959].

Results of the analyses of variance based on the above assumptions are given in table II in the usual format. There, the three basic phonemes C_1, V, and C_2, each of which entails a separate analysis program, are presented separately. Since the critical value of the F-ratio for any level of significance is always greater than unity and since the values of the F-ratios for hypothesis (3) in table II are all less than unity, it must be concluded that this hypothesis cannot be rejected at any level, that is, no significant interaction exists, and the variations observed for "interaction" are attributable to sampling fluctuations.

For hypotheses (1) and (2), the critical value of the F-ratio for the 0.005 level is 1.91 and, since the F-ratios for these hypotheses in all three tests are substantially greater than this, the hypotheses can be rejected with 99.5% confidence. That is, the influence of C_1 and C_2 upon all durations in the triples must be recognized.

PHONEME	SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F RATIO
C_1	TOTAL	4917190	1586	----	
	C_1	2673050	22	121502	86.8
	C_2	143514	22	6523	4.7
	INTERACTION	621160	484	1283	0.9
	RESIDUAL	1479900	1058	1399	----
V	TOTAL	4988880	1586	----	
	C_1	456512	22	20750	22.5
	C_2	3182850	22	144675	156.7
	INTERACTION	373502	484	772	0.8
	RESIDUAL	976457	1058	923	----
C_2	TOTAL	8997140	1586	----	
	C_1	293372	22	13335	5.7
	C_2	5174470	22	235203	101.2
	INTERACTION	1068900	484	2208	0.9
	RESIDUAL	2461030	1058	2326	----

Table II. Analysis of variance results for all three phonemes. All CVC triples including silence are deleted, yielding data in 23×23 matrices.

The realization that consonantal identity does influence durations of consonants located two phonemes away is notable. The situation parallels the interactions observed by ÖHMAN [1966] in his report on VCV formant patterns in which the formants for one vowel were influenced by the phonemic identity of the trans-consonantal vowel. The F-ratio results further show that the durations of single phonemes are not simple functions of the identities of those phonemes, but in addition are functions of their phonemic neighborhoods. To what extent, if any, more distant phonemes might have an effect upon durations, remains to be studied. The relatively small, though extremely significant, values of the F-ratios for the influence of C_1 upon C_2 and of C_2 upon C_1 suggests that the effect of neighboring phonemes decreases very rapidly with distance. Also, it is conceivable that influence would be greater within syllables (as in this study) than across syllables. Verification of these ideas lies in future research.

Since the interaction elements c^p_{ij} are non-significant, equation (1) now becomes

$$D^p_{ij/k} = \mu^p + a^p_i + b^p_j + e^p_{ij/k} \qquad (2)$$

and the ideal representation for the cell means μ^p_{ij} can be given as

$$\mu^p_{ij} = \mu^p + a^p_i + b^p_j \ . \qquad (3)$$

Duration Estimation

Equation (2) can serve not only as a model for the data, but can also provide a method for estimating further durations by deriving parameters from existing data. Since the interaction term c^p_{ij} has been discarded, a value μ^p, a set of 23 a^p_i and 23 b^p_j, can all be determined from one day's data. These 47 parameters may be used to determine the elements $e^p_{ij/k}$ for a given day k since $\hat{D}^p_{ij/k}$, the best estimate of specific duration $D_{ij/k}$, may be expressed

$$\hat{D}^p_{ij/k} = \mu^p + a^p_i + b^p_i, \qquad (4)$$

leading to the relationship

$$e^p_{ij/k} = \hat{D}^p_{ij/k} - D^p_{ij/k} \cdot \qquad (5)$$

Elements $|e^p_{ij/k}|$ may then be averaged over all 529 cells in order to compare the mean deviations for day k with other days and also with other phonemes; the absolute value of $e^p_{ij/k}$ is necessary to prevent means of zero.

Alternate estimates of $D^p_{ij/k}$ may be constructed by utilizing fewer of the 47 parameters. Just μ^p and the 23 a^p_i may be used, or just μ^p and the 23 b^p_j, or simply the mean μ^p. None of these estimators may be considered

mathematically "best estimators," but in certain cases they are approximately as good.

Table III presents the mean deviations obtained for each of the days over each of the phonemes C_1, V, and C_2 under four conditions of duration estimation: 1) μ^p only, 2) μ^p with 23 C_1 parameters, 3) μ^p with 23 C_2 parameters, and 4) μ^p with all 46 C_1 and C_2 parameters. The deviations clearly vary substantially from one to another. As could be expected, the estimation of C_1 using $\mu^p + C_2$ is little better than using μ^p alone; also the estimation using $\mu^p + C_1 + C_2$ is only a slight improvement over $\mu^p + C_1$. In all cases, though, reductions of deviations are effected through the introduction of the C_2 parameters. A parallel statement may be made for C_2 estimation: use of the 23 C_1 parameters produces slight, but consistent, reductions in the mean deviations. When utilized, the C_2 parameters for the vowel bring about a much greater reduction of deviation than do the C_1 parameters; nevertheless, C_1 values make contributions. Generalizing beyond the observation that the longer C_2 durations have considerably greater deviations, one could expect that within a phoneme group the day showing the greatest mean duration would also have the largest mean deviation, and conversely, that the day with the shortest duration would possess the smallest deviation. However, that is not the situation for any of the three phonemes' positions; other factors

		METHOD OF DURATION ESTIMATION			
	DAY	$\bar{\mu}$	$\bar{\mu}+C_1$	$\bar{\mu}+C_2$	$\bar{\mu}+C_1+C_2$
C_1	1	43.7	27.5	42.5	25.9
	2	42.3	23.7	41.7	21.6
	3	44.0	25.1	43.0	24.0
V	1	37.2	35.1	20.6	15.4
	2	41.7	40.1	21.8	16.7
	3	47.2	44.4	23.4	16.4
C_2	1	56.4	53.6	36.4	30.8
	2	67.5	64.6	38.5	32.9
	3	58.1	57.2	34.6	32.3

Table III. Mean deviations in milliseconds of data durations from duration estimations according to four methods: from the mean $\bar{\mu}$ only, from the mean and 23 C_1 parameters, from the mean and 23 C_2 parameters, and from the mean and 46 C_1 and C_2 parameters. Parameters derived from a given data day are used as estimators for that day only. Triples with /#/ are excluded.

are apparently entering in. Nonetheless, it can be seen that various degrees of accuracy may be achieved by using different parametric groups in duration estimation.

Use of the word "estimation" in the preceding discussion is possibly misleading. All 47 parameters used to estimate durations on a given day are derived from the durations they are estimating. For the purposes of this report, this procedure leads to the best possible set of parameters for describing the data from that single day. If the 47 parameters are instead derived from the data of one day and then used to estimate durations from another, a much better test of the ability to predict possible future durations is obtained. Accordingly, day 1 can be used to estimate days 2 and 3; day 2, to estimate days 1 and 3; and day 3, for days 1 and 2. In table IV the averaged mean deviations obtained in this manner are presented along with the averaged mean deviations resulting from one day's use as estimator of its own durations, just presented on a day-by-day basis in table III. In all comparisons, the deviations resulting from the days used as their own estimation source--labeled "DAY ON SELF" in the table--are smaller than those obtained from estimating one with another's parameters, labeled "DAY ON OTHERS" in the table. Note that the amount of reduction obtained by using the 46 C_1 and C_2 parameters in addition to the mean for estimation is markedly greater

		METHOD OF DURATION ESTIMATION			
		$\bar{\mu}$	$\bar{\mu}+C_1$	$\bar{\mu}+C_2$	$\bar{\mu}+C_1+C_2$
C_1	DAY ON SELF	43.3	25.4	42.4	23.9
	DAY ON OTHERS	47.0	36.0	47.1	36.1
V	DAY ON SELF	42.1	39.9	21.9	16.2
	DAY ON OTHERS	47.7	46.7	33.3	31.5
C_2	DAY ON SELF	60.7	58.5	36.5	32.0
	DAY ON OTHERS	61.3	62.9	44.2	46.4

Table IV. Mean deviations in milliseconds of data durations from duration estimation according to four methods: from the mean $\bar{\mu}$ only, from the mean and 23 C_1 parameters, from the mean and 23 C_2 parameters, and from the mean and 46 C_1 and C_2 parameters. Means and parameters derived from one data day are used to estimate durations for that day or for the other two days. Triples with /#/ are excluded.

for the days onto themselves than when they are used upon each other: for C_1, V, and C_2 respectively, the reductions are 45%, 62%, and 47% for the days onto themselves, but the respective reductions are only 23%, 34%, and 24% when the days estimate one another. Apparently, many characteristics of durations existing in just one data day are quite different from those in other days, in spite of the large numbers of measurements comprising each group of 47 parameters. One additional point of interest exists in table IV, in the two consonant sections. In C_1, when the data day estimates its own durations, the estimates formed with μ^p and the 23 C_2 parameters produce a smaller mean, but when the data day values are used to estimate other durations, the mean deviation rises slightly under the same comparison. Likewise, the addition of the 23 C_2 parameters to the μ^p + 23 C_1 parameter group effects a decline in mean deviation for the data day onto itself, but an increase when the data day is used on other days. Parallel statements may be made for C_2: use of the 23 C_1 parameters results in higher mean deviations when one data day is used to estimate others, but not when used upon itself. The quality of the effect of C_1 upon C_2 and of C_2 upon C_1 must be very subtle and may indeed vary in quality from one recording session to another. The alternative is that over just one data day, the effect of consonant identity upon durations so overwhelmed the weak effects of the other consonants, that spurious results were obtained. Since the analysis of variance demonstrates that

C_1 effects upon C_2 and C_2 effects upon C_1 do exist, it might be that the gross variations due to the consonant identity are not sufficiently averaged out in groups of 23 durations, but that they are in groups of 69 durations, permitting the subtler effects to emerge.

Estimation and prediction of phoneme durations can be improved further if one considers the rate of articulation. Obviously, if the speaker were to produce additional CVC utterances, but much more rapidly or slowly than he did, the value of the existing original data as a basis of duration prediction would be impaired, and the error of estimation would increase. More accurate predictions could be generated if the articulatory rate is known.

When the data were originally produced for this research, the speaker attempted to speak at a normal and relaxed rate. It was hoped that the three data sets would have comparable articulatory rates. Indeed, the mean utterance durations are 618 msec, 635 msec, and 690 msec for the three days, respectively; the maximum difference is less than 12%. Nevertheless, differences do exist from one day to another and when one utilizes values from one day as a basis for estimating those of another, these differences influence the resulting errors.

In order to test compensation for the rate of articulation in duration estimation, a series of computer programs with converging parameters were applied to the data. Results

show that a multiplicative model is (almost precisely) the most effective method. Mathematically, if one considers day k to be the source of mean μ^p_k with parameters $a^p_{1/k}$, $a^p_{2/k}$, $\cdots$, $a^p_{23/k}$, $b^p_{1/k}$, $b^p_{2/k}$, $\cdots$, $b^p_{23/k}$, and one wishes to estimate durations of day m whose known mean is μ^p_m, any duration may assume the form:

$$D^p_{ij/m} = \frac{\bar{\mu}^p_m}{\bar{\mu}^p_k} (\mu^p_k + a^p_{i/k} + b^p_{j/k}) + \bar{e}^p_{ij/m} \quad . \tag{6}$$

Rate of articulation for the phoneme (C_1, V, or C_2) under consideration can be expressed as the reciprocal of the mean duration of day m, or $1/\mu^p_m$ phonemes per unit time. Accordingly, a duration estimate constructed from a source data day may be transformed into a target day by multiplying that estimate by the quotient of the rate of articulation of the source day and that of the target day.

Results of duration estimation between pairs of data days, with the compensation applied to rate of articulation according to equation (6), are presented in table V. All deviations are averages of six separate day-to-others combinations; thereby, no data days are used upon themselves. For comparison, the day-to-others results presented previously in table IV are repeated here in the lines "NO ADJ," standing for "no adjustment of rate." The values in the other rows labeled "RATE ADJ" are the new mean deviations for the rate-adjusted deviations. Note that the deviations from a

		METHOD OF DURATION ESTIMATION			
		$\bar{\mu}$	$\bar{\mu}+C_1$	$\bar{\mu}+C_2$	$\bar{\mu}+C_1+C_2$
C_1	RATE ADJ	43.3	30.6	43.4	30.9
	NO ADJ	47.0	36.0	47.1	36.1
V	RATE ADJ	42.1	40.7	23.7	20.4
	NO ADJ	47.7	46.7	33.3	31.5
C_2	RATE ADJ	60.7	62.2	42.7	45.0
	NO ADJ	61.3	62.9	44.2	46.4

Table V. Mean deviations in milliseconds of data durations from rate adjusted and non rate adjusted estimations according to the four methods previously discussed. Triples with /#/ are excluded.

single mean under the rate-adjusted model become the same as those in table IV for data days onto themselves; this is because the quotient μ_m/μ_k in equation (6) effectively transforms the mean of day k into that of day m. It may also be observed that the addition of trans-vowel consonantal parameters still produces an increase rather than decrease in mean deviations.

The improvement of duration estimation by rate adjustment affects the three phonemes very differently. Considering only the 48-parameter method of estimation (the two day means are two of the 48 parameters), C_1 mean deviation is reduced 14% by rate adjustment; for V, improvement is 35%; but for C_2, the deviation is reduced only 3%. Conceivably, the vowel has a higher sensitivity to change of rate on a given day, manifested by greater variation from one to another; possibly also, its durations are more consistent within a given day, while those of the consonants show more variation within one day. Whatever the reason, the vowel mean deviation is cut to less than two-thirds its former value.

The principal results found in table V are the minimum rate-adjusted deviations. Using 25 parameters, the C_1 durations were predicted with a mean accuracy of 30.6 msec; the C_2 durations, with an accuracy of 42.7 msec; and using 48 parameters, V durations were predicted with a mean error of 20.4 msec.

The utility of this rate-adjusted model is limited by its mathematics. The requirement that the rate of articulation of an unknown duration be known can be problematic in areas such as speech recognition. On the other hand, in other areas such as speech synthesis, the entire system of adjustment of rate of articulation may prove to be a convenient method of control. The extent to which the duration values can be expected to remain within the range of a multiplicative model (equation 6) for situations of very slow or rapid speech remains to be seen. Admittedly, the range of variation examined in this study is inadequate for broad generalizations about alternate rates of articulation. Further study of other, more extreme, examples of speech will provide the information.

In the remainder of this section, consideration will be given to further estimation of phoneme durations employing parameter groups of various sizes and constructions.

Results have already been presented in the section for deviation of durations from estimated values for parametric groups of different sizes. Considering only the cases pertaining to rate-adjusted estimation, the following parametric groupings have been seen: 1) two parameters--the means of the source and target days; 2) 25 parameters--the source and target day means plus 23 C_1 parameters; 3) 25 parameters--source and target day means plus 23 C_2 parameters; 4) 48 parameters--the two-day means plus 46 C_1 and C_2 parameters.

Several other alternative groups are possible. The simplest is a set of 531 parameters consisting of two means and 529 cell values for the source day. This has the apparent shortcoming that each cell value is acting as an estimator by itself since it is not averaged in with other durations; its random error $e_{ij/k}$ is included in the estimation, and will result in higher variance of results. Probably, if several groups of data days were averaged, the 531 parameters, while indeed ponderous in size, would prove to be considerably more accurate.

One other method of estimating durations is to consider only the voicing characteristics of C_1 and C_2. Models could be based on: 1) four parameters--the two means plus C_1 voicing means; 2) four parameters--two means plus two C_2 voicing values; and 3) six parameters--the means plus four C_1 and C_2 voicing means.

A final model investigated here is one involving both voicing characteristic and manner of articulation. Each C_1 or C_2 may be described as falling into one of eight categories: sonorant, voiceless fricative, voiced fricative, voiceless sibilant, voiced sibilant, voiceless plosive, voiced plosive, and nasal. Accordingly, models may be constructed similarly to previous ones, but containing now either 10 or 18 parameters.

All the data (/#/ again excluded) were tested for mean deviations from duration estimations derived from these

seven new models. Table VI presents the results averaged in two ways: the data days used as their own estimators, and the data days serving as source estimators for other days. A graphic presentation of these results is given in figure 16. The solid curves in the figure represent the descriptive models developed from a given set of data and used to describe that same set of data; the broken curves represent predictive models derived from one data set, but used to predict the durations of different data sets. Here also for the figure, when the choice of two deviations arose for a given number of parameters, the lower value was selected.

A number of generalizations may be made from the table and accompanying figure: 1) with an increasingly large parametric set, the returns in prediction accuracy of data days on each other diminish and then reverse; 2) there is great prediction stability in situations with few parameters produced by averaging over large numbers of data durations, evidenced by the small difference which exists between low parameter models in figure 16; 3) there is a marked increase in deviation in C_2 when C_1 parameters are introduced, and a slight one in C_1 deviations upon introducing C_2 parameters; likely, if the parameters had been derived from larger data sources, this increase would not exist; and 4) there appear to be minimal levels of mean errors $e_{ij/k}$ where the "DAY TO OTHERS" curves are leveling out--about 20 msec for V, 30 msec for C_1, and 40 msec for C_2.

NUMBER OF PARAMETERS				C_1		V		C_2	
$\bar{\mu}$	C_1	C_2	SUM	DAY TO SELF	DAY TO OTHERS	DAY TO SELF	DAY TO OTHERS	DAY TO SELF	DAY TO OTHERS
2	-	-	2	43.3	43.3	42.1	42.1	60.7	60.7
2	2	-	4	37.4	37.5	40.7	40.7	60.3	60.9
2	-	2	4	43.3	43.3	34.1	34.1	56.5	57.0
2	2	2	6	37.4	37.6	32.6	32.5	56.0	57.2
2	8	-	10	27.8	30.8	40.3	40.8	59.1	61.9
2	-	8	10	43.1	43.6	25.8	26.8	42.5	45.6
2	8	8	18	27.2	31.1	22.2	24.2	39.6	47.1
2	23	-	25	25.4	30.6	39.9	40.7	58.5	62.2
2	-	23	25	42.4	43.4	21.9	23.7	34.8	42.7
2	23	23	48	23.9	30.9	16.2	20.4	32.0	45.0
2	529		531	0.0	37.3	0.0	23.2	0.0	53.1

Table VI. Mean deviations in milliseconds resulting from estimating the C_1, V, C_2 durations according to 11 different parametric models. Results are separated into two groups depending upon whether the data day used for creating the parameters was used back upon itself or upon other data days. No /#/ is included.

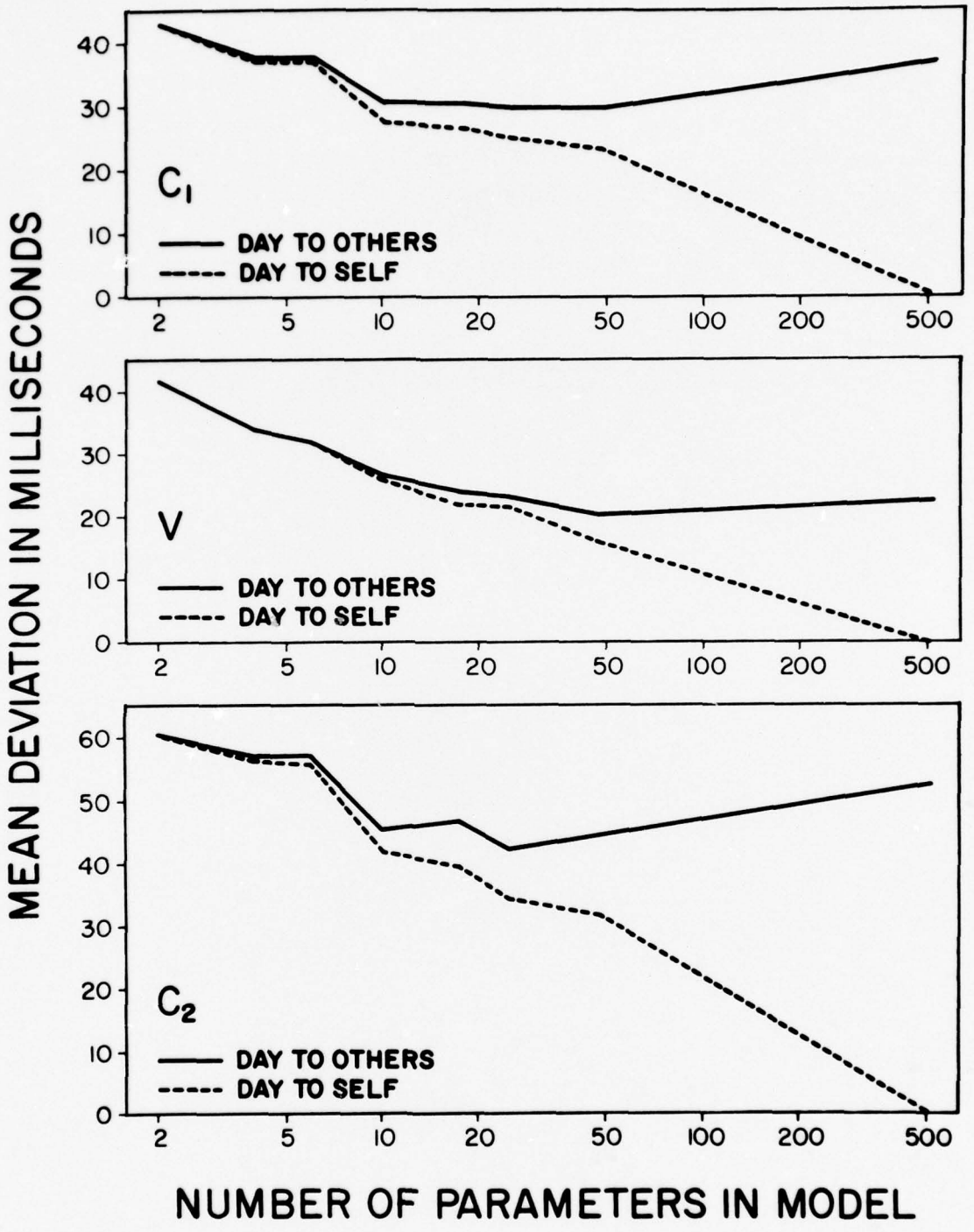


Figure 16. Mean absolute errors of various models for the durations of the initial consonant, the vowel, and the final consonant, plotted according to the number of parameters used.

The area of duration prediction is complex and involves many subtle influences which surface only upon extensive statistical treatment. Nevertheless, when stability is achieved, the interrelationships emerge and may be recognized, expected, controlled, and recreated.

In concluding this section, the following aspects are deemed to be the most important:

1) The analysis of variance demonstrates that both C_1 and C_2 exert statistically significant influences upon all durations within the C_1VC_2 utterances.

2) The second consonant exerts a greater influence upon itself and the contiguous vowel than C_1 does upon itself and V, but the C_2 influence upon C_1 is weaker than that of C_1 on C_2.

3) An interaction effect upon durations is statistically non-significant.

4) Duration estimation through the utilization of various parametric models reveals that estimation accuracy improves with parameter increase up to 10 parameters for C_2, but up to 25 parameters for C_1 and V; and that above that, accuracy decreases.

5) Random fluctuations in the durations appear to be about 30 msec for C_1, 20 msec for V, and abut 40 msec for C_2.

References

Blalock, H. M.: Social statistics (McGraw-Hill, New York 1960).

Broad, D. J., and Fertig, R. H.: Formant-frequency trajectories in selected CVC-syllable nuclei. J. acoust. soc. Amer. 47: 1572-1582 (1970).

Dixon, W. J., and Massey, F. J. Jr.: Introduction to social statistics (McGraw-Hill, New York 1951).

Koenig, W., Dunn, H. K., and Lacy, L. Y.: The sound spectrograph. J. acoust. soc. Amer. 18: 19-49 (1946).

Lisker, L., and Abramson, A. S.: Some effects of context on voice onset time in English stops. Language and Speech 10: 1-28 (1967).

Öhman, S. E. G.: Coarticulation in VCV utterances: spectrographic measurements. J. acoust. soc. Amer. 39: 151-168 (1966).

Peterson, G. E., and Lehiste, I.: Duration of syllable nuclei in English. J. acoust. soc. Amer. 32: 693-703 (1960).

Scheffé, H.: The analysis of variance (Wiley, New York 1959).

Shoup, J. E.: The phonemic interpretation of acoustic-phonetic data. Ph.D. dissertation. The University of Michigan 1964).

Potter, R. K., Kopp, G. A. and Green, H. C.: Visible Speech (Van Nostrand, New York 1947); (Dover, New York 1966).

Appendix I

Segmentation Criteria

Initial #. In two-thirds of the cases in which /l/ was preceded by silence, there was a breathy period of up to 80 milliseconds duration. During this period the formants were generally recognizable and similar in appearance to an /h/ which is immediately adjacent to a vowel. The onset of voicing was always strongly evident in the first formant, however, and usually occurred simultaneously in the first three formants, and segmentation was performed here.

Initial /w/. Initial /w/ consists of a steady-state followed by a rapid rise in the second formant toward the vowel. Although Peterson and Lehiste (1960) prefer to segment /w/ where the slope of the second formant becomes positive, it was decided here to segment at the reflex point, i.e., the maximum rate of increase of the second formant, principally because of the absence or extreme weakness of the third formant in this initial region in virtually all cases.

Initial /l/. The onset of the vowel is almost always demarked by a shift in formant rate of change and intensity in the first and third formants and, most particularly, in the second formant. Segmentation was performed at this onset.

Initial /r/. As with the /w/, initial /r/ was segmented at the point of maximum rate of increase of the second formant. If the rate remained maximal for a substantial interval, as

happened several times, segmentation was performed at the middle of this interval. This avoided the ambiguity of the Peterson-Lehiste criteria which depend upon friction, third formant movements, and change from steady-state to "onglide."

Initial /j/. For initial /j/, Peterson and Lehiste describe "a rapid dip in frequency" of the third formant, the lowest value of which they used for segmentation of their data. Potter, Kopp & Green (1947, p. 214) also describe such a dip, and note its occurrence "frequently" for /ja/ and /jɑ/, but fail to mention it elsewhere. In this study dealing only with /jɪ/, the dip did not occur at all. In general, however, there was a strong increase in intensity of the second formant which was used for segmentation. This occurred where the third formant had already left the /j/ steady-state and was moving downward toward the /ɪ/.

Initial /ʍ/. The problem here was whether or not to use the onset of voicing as the segmentation point. Voicing commenced over a large range of positions, from as far as 80 msec prior to the beginning of the rise of the second formant, to approximately the point of maximum rate of increase of the second formant. It was decided to segment the /ʍ/ at the same position as the /w/; that is, at the point of maximum rate of increase of the second formant.

Initial /f/. The obvious place of segmentation was the onset of voicing, which was used. Indeed, there was no

consistent formant variation near the voicing onset, and the formants remained nearly unchanging into the /ɪ/.

Initial /v/. Initial /v/ was segmented either at the strong increase of intensity of at least one (and usually more) of the first three formants or at the slight rise in frequency of the first formant. When more than one of the above occurred, they were simultaneous. It might be noted that Peterson and Lehiste used the cessation of friction for segmenting initial voiced fricatives, and in the data spectrograms, friction was indeed observed in all cases. This friction was, however, frequently very weak and could have continued unobserved in many cases during the very prominent vowel sound. Whether or not that did indeed occur, the onset of strong vowel voicing was used for segmenting.

Initial /θ/. As with initial /f/, the onset of voicing was used for segmentation. A characteristic which appeared to be peculiar to the /θ/-/ɪ/ combination was a slight burst of noise about 25-70 msec prior to the onset of voicing. This burst, which was observed in over half of the data, was almost always followed by a period of less intense friction than that before. It might be conjectured that this burst is created by a mechanism similar to that of plosives, but in this case with an incomplete closure. Its absence would therefore indicate a more open articulation.

Initial /ð/. The initial /ð/ is extremely similar to initial /v/ and was segmented at the point of onset of strong

formant patterns for at least one of the formants, or, alternately, the point of rise of the first formant.

Initial /h/. Since there is little or no energy in the /h/ below 2000 Hz, the onset of voicing of the first formant provided a definite location for segmentation. Voicing in the higher formants appeared to start simultaneously, but was frequently difficult to distinguish from the heavy friction of the /h/.

Initial /s/. Onset of voicing of the first formant provided the clearest place for segmentation.

Initial /z/. As with the voiced fricatives, initial /z/ was segmented at the place of increase of intensity of the formants. One characteristic which appears to be peculiar to both the /z/ and /ʒ/ in initial position is the occurrence of a gentle drop and rise in the first formant. This seems to take place in the last half of the sibilants and could also be used to signal the start of the vowel.

Initial /ʃ/. Initial /ʃ/ was very clearly segmentable at the point of onset of voicing of the first formant.

Initial /ʒ/. These data are similar to the initial /z/, except for slight shifts in formants and a much higher intensity in the /ʒ/. This higher intensity tended to obscure any rise in intensity of the third formant which was helpful in segmenting the initial /z/. The increase in intensity and rise in location of the first formant were typically rather gradual and therefore not very useful either.

This left the second formant whose rise in intensity was generally used for segmentation, with the occasional aid of changes in the first and third.

Initial /p/. Initial /p/ was segmented at the onset of voicing of the vowel. It is significant that all duration values for initial plosives do not include the period of closure; voiceless plosive values represent the duration of aspiration, that is, the interval between the burst and voicing onset, while voiced plosive values represent the duration of the voice bar.

Initial /b/. Peterson and Lehiste segmented initial voiced plosives at the center of the spike which resulted from release of the pressure built up during closure. They also segmented initial voiceless plosives at two different locations, one at the center of the spike and one at the onset of voicing. Since we decided in this study to consider only voiced segments as possible vowels, both voiceless and voiced plosives were segmented at the onset of full formants. Voicing began prior to the spike (and hence prior to the full formant pattern) in all of the initial /b/ phoneme samples except for one in which they occurred simultaneously.

Initial /t/. Initial /t/ was segmented at the onset of voicing after the aspiration.

Initial /d/. Here segmentation was performed after the initial spike and at the onset of full formants. In 11 of the 72 initial /d/ phoneme samples, there was no voicing prior to the burst.

Initial /k/. As for initial /p/ and /t/, segmentation was done at the voicing onset.

Initial /g/. Initial /g/ was segmented as were initial /b/ and /d/, at the onset of full formants after the spike. Voicing onset was delayed in 20 of the 72 samples until at least the spike location and, in many cases, it was concurrent with the formant pattern onset.

Initial /m/. Initial /m/ was easily distinguished from the vowel by an abrupt change in the location or slope, or both, of the first three formants. In these data, the /m/ was not characterized by steady formant patterns; in every case, there were shifts of location of at least one of the formants. This reflects shifting resonating cavities during the oral closure and velar opening, preparatory to formation of the vowel from a neutral position at onset.

Initial /n/. As with initial /m/, initial /n/ is sharply differentiated from the vowel by obvious changes in formant location or slope, and, in addition, often by structure and intensity as well. The second /ɪ/ formant in particular appeared strongly and fully in an area previously very low or lacking in energy on the spectrogram. Segmentation was performed at the location of these changes.

Initial /ŋ/. Here, due to areas of resonance in the /ŋ/ which tend to correspond to the first and second formants of /ɪ/, it is difficult if not impossible to use these for segmentation. The third formant of the vowel, however, commenced

abruptly in an area of non-resonance of the /ŋ/, so this consistent feature was used for segmenting.

Final #. Without exception, final /#/ was characterized by a period of breathiness which continued the formant pattern after the voicing had died out. Indeed, it was often present simultaneously with the voiced /l/ during the last several pulses. Segmentation was performed at the last obvious pulse.

Final /w/. Final /w/ presented no problems as it was segmented the same way as initial /w/, that is, at the point of maximum rate of change (this time, decrease) of the second formant. The tendency of the "unnatural" terminal /w/ to return to a relaxed /ə/ position was difficult for the speaker to control consistently. In 12 of the 72 recordings of final /w/, the second formant had again passed a point of maximum increase rate, at which the remaining fragmentary /ə/ was segmented and not included in the /w/ duration.

Final /l/. Peterson and Lehiste observed in many final /l/ situations a rapid drop and rise of the third formant, at which point, if present, they performed their segmentation. Otherwise, they used the leveling of the fundamental frequency as a determining factor. In this study, 49 of the 72 contained the third formant drop and rise. In order to have one universal measure, and to be as consistent in segmentation as possible, it was decided to segment final /l/ at the point of maximum change of the second formant. This point generally occurred one or two voicing pulses after the dip in the third formant where this dip was present.

Final /r/. Final /r/ was handled the same as initial /r/. Segmentation was performed at the point of maximum change of slope of the second formant; or, if there was a sizable interval with such slope, it was performed at the middle of that section.

Final /j/. Final /j/ was perhaps the most difficult consonant to segment. There was usually, but not always, a rise in both the third and second formants concurrent with a slight drop in the first formant. Often the second and third formants rose at a steady rate from the initial consonant continuously through the vowel on to the high point of /j/. One characteristic in virtually all 72 spectrograms was a noisy area of reduced formant intensity. If this was present, the beginning of this area was used for segmentation. In the cases which did not possess this noise, a point just after the place where the second and third formants started to rise was used.

Final /ɱ/. As could be expected, the termination of voicing took place at greatly varied positions between the /ɪ/ and the non-English final /ɱ/. For this reason, and in order to be consistent with the initial /ɱ/ and other segmentation, segmentation was done at the point of maximum slope decrease of the second formant. A comparison of the end of voicing with the place of maximum F_2 slope, shows that 32% of the data had these 2 locations within approximately 10

msec of each other, 22% had the voicing end before, and 46% had the voicing end after. This is the one exception to the voiced vowel criterion initially established for voiceless consonants.

Final /f/. Unlike initial /f/ where the voiceless formants of the /f/ lead horizontally into the vowel, final /f/ is characterized by sloping formants in virtually all cases. Here we find the first formant rising while the second and third fall. But as with initial /f/, segmentation is performed at the termination of the vowel voicing.

Final /v/. Peterson and Lehiste used the "rapid decrease" of energy to segment final /v/'s from preceding vowels. In this study the spectrograms showed a gradual onset of friction and gradual offset of voicing. It was noted, however, that the first formant continued strongly into the /v/; the second not as far as the first, and the third again not as far as the second. In general, there appears to be a place where the steadily dropping second and third formants tend to level out. It is at approximately this same place where the third formant dies out and the second drops in intensity. It was determined to segment at this point, which possibly corresponds to the intensity drop used by Peterson and Lehiste.

Final /θ/. In virtually all spectrograms in this category frictional energy existed prior to the end of the voicing by several pulses. Due to the frequent drop in energy of the formants at the voicing termination, and in order to be

consistent with initial /θ/ segmentation, it was decided to segment final /θ/ at the end of the voicing. This is different from the decision of Peterson and Lehiste to segment at the onset of friction. As they wrote, "The vowel was considered terminated at the point where the noise pattern began, even though voicing in a few low harmonics continued for a few centiseconds in most cases." In this study it was found that the noise onset was rarely abrupt, but rather gradual in appearance. The end of the voicing, on the contrary, was more readily determined. This was not only true for the strong first formant, but frequently for the second and third in which the voicing could clearly be observed to the end before the friction took over exclusively.

Final /ð/. Final /ð/ is very similar to final /v/ and was treated in the same way. Indeed, the drop in intensity of the formants seems to be even more pronounced in these cases, so this point, being fairly consistent with the bottom of the drop of F_2, was used for segmentation.

Final /h/. As with other final voiceless fricatives, final /h/ was segmented at the termination of voicing. In contrast with the final /θ/ and /f/, however, the voicing in the second and third formant regions appeared to be more thoroughly dominated with noise before the voicing in the first formant ended.

Final /s/. Final /s/, being preceded by a fairly short vowel, was typified by an abrupt onset of noise followed

by a more abrupt offset of voicing within a few centiseconds. The voicing in the higher formants generally continued until the termination of all voicing, at which point segmentation was performed.

Final /z/. Peterson and Lehiste segment final voiced sibilants at the "onset of high frequency energy" observed clearly in their intensity curves. In the absence of these aids, we had to rely on spectrograms. In the majority there exist points at which at least one and usually two of the formants greatly decrease in intensity, concurrently with a leveling out of the drop of the second formant. As the drop in intensity was abrupt, it was used for segmentation when possible. When the drop was unclear or gradual, segmentation was performed where the second formant ceased dropping sharply.

Final /ʃ/. The transition from the vowel to the sibilant /ʃ/ is typified by a distinct intermediate stage lasting about 20-40 milliseconds. In this stage we find a great reduction in intensity of the higher formants, an introduction of light noise, and an undisturbed continuation of the first formant. At the end of this stage, the formants abruptly die and the light noise is replaced with the high-frequency, high-intensity sibilant noise. The formant directions continue their prior course during this intermediate stage. Because of this, and in order to maintain consistency with other segmentations, the final /ʃ/ was segmented at the termination of voicing.

Final /ʒ/. Final /ʒ/ is one of the most difficult consonants to segment. Unlike the other voiced sibilants and fricatives, final /ʒ/ contained a formant positioning extremely similar to that of the preceding vowel. Here there is no distinct drop of F_2 to aid in segmentation, although there are frequently drifts, upward and downward, of both the second and third formants. The problem is further compounded by the extreme durations of both the vowel and the consonant /ʒ/ and, accordingly, of their subtle transition. There is, however, in most spectrograms, a place where the intensity of the second and third formants decidedly decreases. This invariably occurs just after the onset of light noise. When this occurs, segmentation was performed at this place. If, however, the formants continued strongly into the sibilant noise, segmentation was performed at a location a few centiseconds after the commencement of the noise. This created as much consistency within the group of final /ʒ/ spectrograms and preserved as much similarity to the other segmentation procedures as possible.

Final /p/. Final /p/ provides no problems and segmentation was performed at the cessation of all formants, which always occurs abruptly.

Final /b/. The segmentation of final /b/ is accurately positioned at the abrupt end of the higher formants concurrent with a distinct drop in the first formant. The

third formant tends to weaken considerably in intensity before its termination, but is always traceable to the end.

Final /t/. Final /t/, like final /p/, is easily segmented at the simultaneous cessation of all formants.

Final /d/. As with final /b/, segmentation was performed at the abrupt end of F_2 and F_3 and the drop of F_1.

Final /k/. Segmentation for final /k/ was executed at the point where all formants abruptly ended.

Final /g/. Here, as with the other final voiced plosives, segmentation was performed at the end of F_2 and F_3. A curious phenomenon peculiar to final /g/ is the near or apparent coincidence of the second and third formants at the end of the vowel. In many spectrograms it is extremely difficult to imagine that the formant centers do anything but contact at the very point at which they cease to exist. In the other spectrograms they are moving on convergent paths at the time they are cut off, as if they would soon contact in the region of oral closure.

Final /m/. The very abrupt and simultaneous shift of location of the first formant and shift of direction of the second formant were used as the location for segmenting the final /m/. Of interest, though, is the third formant which in general dipped slightly before gradually rising through the first part of the /m/ to a steady formant position. The low point of this dip sometimes coincided with the abrupt F_1 and F_2 changes, but more typically took place a few

centiseconds before. Nasalization of the vowel was apparent in the last third of the vowel in most cases, but failed to interfere with segmentation.

Final /n/. For the final /n/, segmentation was performed at the point where the formants all abruptly changed direction, location, or intensity. Unlike the case for the /m/, the third formant here moves fairly directly to a point at the end of the vowel where it continues throughout the /n/ in a steady formant position. Vowel nasalization, while apparent, did not obscure any of the vowel formants or interfere with segmentation.

Final /ŋ/. As with the other nasals, final /ŋ/ was easily segmented. Here, the first formant frequently continued with little or no change from vowel to consonant, so segmentation clues came from the abrupt changes in the second, third, and frequently fourth formants. Nasalization, while apparent in most cases, in no way obscured the transition to prevent accurate segmentation.

Appendix II

The following tables of mean durations and standard deviations are offered for individuals who may be interested in explicit values derived from the data. They consist of three basic mean and standard deviation groups: those derived from C_1 durations, from V durations, and then from C_2 durations. Only the C_1VC_2 triples are included; the C_1V and VC_2 phoneme pairs have been deleted. Each of these three basic mean and standard deviation groups has been broken down into two subgroups, those derived while holding C_1's constant and ranging over the 23 C_2 possibilities, and those which hold C_2's constant and range over the 23 C_1's. Each of these subgroups has in turn been split into each of the three component data-days. Accordingly, each value given in the following list represents in a mean or standard deviation derived from 23 measured durations.

C_1 DURATIONS

C_1	DAY 1		DAY 2		DAY 3	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
w	149.0	30.5	134.0	25.9	171.1	27.8
l	137.7	30.6	142.0	24.9	153.1	28.9
r	112.6	17.2	104.0	16.3	124.7	24.1
j	132.4	29.6	116.2	29.0	135.8	31.1
ɱ	182.5	39.7	240.7	39.7	253.4	33.0
f	220.7	32.6	207.7	32.5	232.7	34.5
v	167.0	22.7	158.1	36.9	168.4	27.5
θ	235.3	28.6	208.4	41.3	258.7	28.4
ð	174.9	45.7	153.4	35.5	179.9	32.0
h	208.9	40.7	220.7	33.4	216.6	48.7
s	218.5	32.8	223.7	33.9	254.8	26.3
z	207.5	33.7	148.1	19.9	177.7	28.1
ʃ	244.9	24.0	219.0	26.8	261.6	29.5
ʒ	207.9	38.1	178.7	31.9	203.6	24.3
p	155.4	38.5	101.3	27.5	151.9	38.3
b	137.6	27.0	122.0	23.5	130.6	34.2
t	158.4	26.7	121.3	15.6	166.9	16.1
d	92.8	52.4	110.4	39.7	141.7	32.2
k	180.0	33.8	134.6	20.6	187.2	26.2
g	91.4	65.2	103.9	45.5	127.6	40.2
m	156.7	36.3	128.9	23.2	166.5	27.2
n	148.3	32.1	124.4	28.1	168.1	35.1
ŋ	176.4	35.4	144.8	36.3	227.4	48.3

C_1 DURATIONS

C_2	DAY 1		DAY 2		DAY 3	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
w	180.7	51.9	138.8	51.6	170.0	44.1
l	190.6	41.6	170.0	46.1	202.2	44.9
r	197.2	49.3	172.6	53.5	204.5	50.6
j	160.1	53.6	143.3	56.5	179.4	43.8
ʍ	161.8	57.9	140.5	52.5	172.8	51.6
f	191.8	54.5	156.9	60.7	187.8	55.8
v	164.9	49.9	157.7	49.4	183.9	61.7
θ	156.8	58.7	161.1	66.5	183.6	70.6
ð	158.4	47.3	149.2	50.2	159.1	61.7
h	171.2	51.7	133.3	52.1	179.1	54.6
s	175.1	77.3	169.0	62.1	187.1	50.3
z	176.1	64.3	171.0	48.4	199.0	50.8
ʃ	166.6	49.9	173.6	51.8	187.1	53.7
ʒ	164.7	58.5	163.0	50.6	176.5	66.0
p	180.4	52.6	143.3	35.4	191.4	59.5
b	150.8	49.1	157.0	42.3	188.1	46.5
t	170.8	47.6	153.5	53.5	187.8	54.3
d	169.3	44.4	143.9	53.0	200.5	62.8
k	167.1	51.9	138.0	45.8	184.1	50.0
g	154.8	68.4	137.0	52.0	165.3	42.1
m	163.3	44.9	155.9	54.9	184.9	54.4
n	151.0	46.6	154.8	35.6	192.7	42.1
ŋ	173.9	46.2	162.8	56.5	193.1	36.3

V DURATIONS

C_1	DAY 1		DAY 2		DAY 3	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
w	130.6	31.6	146.7	41.3	167.3	46.9
l	163.9	43.2	173.3	51.0	194.4	49.4
r	162.1	39.3	176.7	41.0	212.1	55.3
j	162.2	42.3	198.2	44.7	224.7	57.2
M	136.7	29.0	147.3	41.6	177.8	55.4
f	139.0	36.0	160.7	42.8	175.4	46.9
v	157.8	36.0	175.6	38.5	221.5	55.9
θ	136.8	38.0	157.8	47.6	183.7	56.5
ð	158.5	40.5	185.1	52.4	211.7	54.6
h	145.9	48.3	153.7	45.4	174.4	46.4
s	129.3	42.0	150.1	45.0	169.3	49.3
z	156.3	51.2	181.3	49.1	196.1	51.9
ʃ	139.5	43.9	153.0	48.4	170.6	51.2
ʒ	153.2	46.6	165.7	47.6	180.7	48.5
p	138.6	40.8	156.3	49.5	175.8	55.1
b	168.7	49.5	197.8	57.3	209.3	63.1
t	141.1	43.1	162.2	54.1	178.9	57.2
d	188.4	58.4	202.7	65.3	216.6	53.0
k	128.7	36.6	161.4	50.7	182.7	56.0
g	169.6	41.0	197.7	64.1	228.7	62.0
m	160.7	49.8	165.8	56.5	193.3	66.8
n	167.8	56.9	172.8	58.0	184.6	46.8
ŋ	198.0	75.6	180.0	66.1	215.3	68.0

V DURATIONS

C_2	DAY 1		DAY 2		DAY 3	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
w	137.0	29.5	137.3	24.9	169.8	25.1
l	143.2	21.8	153.7	19.3	183.2	25.6
r	219.2	29.9	231.3	29.7	265.7	29.2
j	181.2	30.2	186.5	29.2	219.6	43.4
ʍ	145.8	28.1	143.7	23.0	182.6	26.1
f	119.3	17.1	134.1	18.9	150.2	19.2
v	194.1	34.1	225.0	34.1	253.3	34.4
θ	131.1	17.9	148.0	20.6	159.9	23.2
ð	220.7	46.8	233.8	39.3	262.1	38.3
h	157.4	22.9	145.9	19.0	205.1	46.2
s	125.8	13.4	132.0	19.4	149.4	27.0
z	218.3	35.6	260.9	51.1	279.6	44.5
ʃ	123.0	22.1	142.6	19.6	143.6	20.6
ʒ	237.9	57.1	262.1	47.4	295.1	32.1
p	91.4	17.9	96.7	14.8	115.9	17.0
b	143.1	14.1	165.0	24.7	176.5	21.8
t	99.8	18.5	116.0	22.7	134.7	31.5
d	163.3	23.5	183.5	29.8	227.2	34.5
k	95.4	17.7	104.6	30.8	119.0	23.5
g	142.6	21.2	174.5	21.6	203.1	22.4
m	147.8	23.8	177.3	23.6	182.2	23.2
n	158.6	26.6	195.4	26.3	197.5	28.5
ŋ	137.8	22.9	172.2	25.9	169.6	18.3

C_2 DURATIONS

C_1	DAY 1		DAY 2		DAY 3	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
w	294.7	59.3	330.5	65.6	304.9	76.3
l	302.3	64.0	340.8	68.1	313.3	68.3
r	271.9	68.5	328.1	77.5	305.1	55.9
j	313.7	82.2	332.5	75.2	310.3	66.7
ʍ	268.1	62.6	342.1	80.5	298.0	78.8
f	310.1	86.1	326.2	89.0	301.5	74.6
v	323.1	71.3	340.3	79.0	312.5	71.0
θ	282.7	75.1	317.4	95.7	305.7	80.3
ð	289.4	61.6	344.7	63.9	342.2	67.8
h	277.2	63.4	300.1	59.6	288.7	52.6
s	276.6	59.6	324.7	73.5	308.0	56.4
z	279.9	67.6	243.7	56.2	307.0	85.2
ʃ	271.3	43.0	294.3	79.5	294.0	61.0
ʒ	268.3	56.3	282.5	85.2	314.6	65.1
p	263.5	62.6	310.6	84.7	335.6	92.3
b	308.7	62.3	288.7	75.3	313.5	64.3
t	288.5	71.5	298.4	77.9	312.4	67.6
d	312.2	36.4	297.6	77.1	311.2	57.3
k	265.3	41.8	290.1	82.5	288.7	64.2
g	301.7	70.0	301.5	88.0	321.8	70.2
m	328.2	80.3	288.4	75.6	328.6	77.8
n	349.9	59.0	298.0	88.9	326.4	70.5
ŋ	331.5	83.0	303.6	94.0	317.7	87.9

C_2 DURATIONS

C_2	DAY 1		DAY 2		DAY 3	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
w	209.4	27.6	186.6	32.8	227.0	31.3
l	219.5	34.3	225.4	30.0	272.7	45.6
r	202.8	36.5	210.1	36.0	255.7	35.3
j	259.2	46.1	290.6	44.8	251.8	37.3
m	217.5	47.3	223.3	48.9	245.9	46.2
f	286.1	42.5	321.5	52.7	345.4	35.2
v	284.3	48.8	261.3	58.4	275.1	33.5
θ	306.7	42.4	361.4	41.5	340.8	44.9
ð	283.4	52.9	301.7	76.2	279.0	47.9
h	214.3	59.9	227.5	66.6	206.0	53.8
s	384.3	51.0	422.8	55.7	418.7	40.4
z	308.2	36.9	330.5	53.7	300.6	37.0
ʃ	370.3	45.1	427.7	47.8	409.0	42.9
ʒ	309.9	40.5	349.7	45.3	314.4	33.6
p	329.7	69.0	367.8	65.4	359.4	63.7
b	310.2	46.3	324.2	50.1	298.9	37.0
t	366.1	41.3	397.7	41.7	360.8	52.8
d	305.7	39.4	295.6	43.2	289.3	33.9
k	374.6	53.2	389.1	44.5	401.2	50.5
g	310.8	45.7	269.9	46.3	263.9	32.7
m	288.7	46.4	303.7	39.3	345.1	42.7
n	317.8	38.3	305.6	33.5	328.2	47.4
ŋ	319.0	46.9	331.4	35.7	372.8	42.7

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TEMPORAL INTERRELATIONS IN SELECTED ENGLISH CVC UTTERANCES		5. TYPE OF REPORT & PERIOD COVERED Scientific-Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ralph H. Fertig		8. CONTRACT OR GRANT NUMBER(s) F44620-69-C-0078 F44620-74-C-0034 N00014-67-C-0118 N00014-75-C-0483
9. PERFORMING ORGANIZATION NAME AND ADDRESS Speech Communications Research Laboratory, Inc., 800 A Miramonte Drive Santa Barbara, California 93109		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Directorate of Mathematical & Information Sciences, United States Air Force Office of Scientific Research, Bldg. 410, Bolling AFB, Washington, D.C.		12. REPORT DATE May 1976 13. NUMBER OF PAGES 83
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Monograph
SCRL Monograph Number 12
May 1976

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

consonant
prosody
prosodic variability
sound spectrography
speaker variability
speech
speech duration
vowel

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The segment duration in a corpus of 1728 CVC utterances containing the vowel /ɪ/ by a single speaker were measured and analyzed. Measurement of sound spectrograms was facilitated by a computer assisted television display. Various distributions of the durations under different conditions reveal a number of effects on segment durations attributable to consonant identity, consonant voicing characteristics and manners of articulation. Analyses of variance →

Block 11. (continued) Controlling Office Name and Address

Directorate of Mathematical and Information Sciences
United States Air Force Office of Scientific Research
Bldg. 410
Bolling Air Force Base, Washington, D.C.

For contracts F44620-69-C-0078 and F44620-74-C-0034

Department of the Navy
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

For contracts N00014-67-C-0118 and N00014-75-C-0483

Block 20. (continued) Abstract

show that an interaction effect between the two consonants is statistically not significant for the durations of either consonant or of the vowel. Various models for describing and predicting the durations are given together with a discussion of their associated errors. The segmentation criteria are given in some detail as an appendix. A second appendix contains basic means and standard deviations derived from the data.

SCRL MONOGRAPH SERIES

No. 1 Huttar, G. L., *Some Relations Between Emotions and the Prosodic Parameters of Speech*, July, 1967.

No. 2 Houde, R. A., *A Study of Tongue Body Motion During Selected Speech Sounds*, August, 1968.

No. 3 Broad, D. J., *Some Physiological Parameters for Prosodic Description*, October, 1968.

No. 4 Benguerel, A-P and Grunstrom, A. W., *Studies in French Grammar and Phonology*, November, 1968.

No. 5 Markel, J. D., *On the Interrelationships Between a Wave Function Representation and a Formant Model of Speech*, May, 1970.

No. 6 Aurbach, J., *A Phonemic and Phonetic Description of the Speech of Selected Negro Informants of South-Central Los Angeles*, March, 1971.

No. 7 Markel, J. D., *Formant Trajectory Estimation from a Linear Least-Squares Inverse Filter Formulation*, October, 1971.

No. 8 Ishizaka, K. and Matsudaira, M., *Fluid Mechanical Considerations of Vocal Cord Vibration*, April, 1972.

No. 9 Wakita, H., *Estimation of the Vocal Tract Shape by Optimal Inverse Filtering and Acoustic/Articulatory Conversion Methods*, July, 1972.

No. 10 Markel, J. D., Gray, Jr., A. H., Wakita, H., *Linear Prediction of Speech-Theory and Practice*, September, 1973.

No. 11 Earle, M. A., *An Acoustic Phonetic Study of Northern Vietnamese Tones*, June, 1975.

No. 12 Fertig, R. H., *Temporal Interrelations in Selected English CVC Utterances*, May, 1976.

Speech Communications Research Laboratory, Inc.
800-A Miramonte Drive
Santa Barbara, California 93109